

1Hybrid capture data unravels a rapid radiation of pimpliform parasitoid wasps
2(Hymenoptera: Ichneumonidae: Pimpliformes)

3Seraina Klopstein^{1,2}, Barbara Langille³, Tamara Spasojevic^{1,2}, Gavin R. Broad⁴, Steven J. B.

4Cooper^{3,5}, Andrew D. Austin³, Oliver Niehuis^{6,7}

7¹ University of Bern, Institute of Ecology and Evolution, Baltzerstrasse 6, 3012 Bern,
8Switzerland

9² Naturhistorisches Museum Bern, Bernastrasse 15, 3005 Bern, Switzerland

10³ Australian Centre for Evolutionary Biology and Biodiversity, School of Biological Sciences,
11The University of Adelaide, North Terrace, Adelaide, SA 5005, Australia

12⁴ Natural History Museum, Department of Life Sciences, Cromwell Road, London SW7 5BD,
13UK

14⁵ South Australian Museum, Evolutionary Biology Unit, North Terrace, Adelaide, SA 5000,
15Australia

16⁶ Center for Molecular Biodiversity Research, Zoological Research Museum Alexander
17Koenig, Adenauerallee 160, 53113 Bonn, Germany

18⁷ University of Freiburg, Department of Evolutionary Biology and Ecology, Institute for
19Biology I (Zoology), Hauptstraße 1, 79104 Freiburg, Germany (current affiliation)

20

21Running head: Rapid radiation of pimpliform parasitoid wasps

22

23Corresponding author: Seraina Klopstein, Naturhistorisches Museum Bern, Bernastrasse 15,
243005 Bern, Switzerland, Tel. +41 31 350 72 45, E-mail: klopstein@nmbe.ch

25

26Abstract

27The parasitoid wasp family Ichneumonidae is among the most diverse groups of organisms,
28with conservative estimates suggesting that it contains more species than all vertebrates
29together. However, ichneumonids are also among the most severely understudied groups, and
30our understanding of their evolution is hampered by the lack of a robust higher-level
31phylogeny of this group. Based on newly generated transcriptome sequence data, which were
32filtered according to several criteria of phylogenetic informativeness, we developed target
33DNA enrichment baits to capture 93 genes across species of Ichneumonidae. The baits were
34applied to DNA of 55 ichneumonids, with a focus on Pimpliformes, an informal group
35containing nine subfamilies. The target DNA capture efficiency was consistently high across
36the investigated species, including three distantly related outgroups. Phylogenetic trees were
37inferred under maximum likelihood and Bayesian approaches, both at the nucleotide and
38amino acid levels. We found maximum support for the monophyly of Pimpliformes but low
39resolution and very short branches close to its base, strongly suggesting a rapid radiation.
40Different analytical approaches recovered different trees, which precludes robust inferences of
41the evolution of different parasitoid strategies, such as idiobiosis and koinobiosis. Two genera
42and one genus-group were consistently recovered in unexpected parts of the tree, prompting
43changes in their higher-level classification: *Pseudorhyssa* Merrill, currently classified in the
44subfamily Poemeniinae, is transferred to the tribe Delomeristini within Pimplinae, and
45*Hemiphanes* Förster is moved from Orthocentrinae to Cryptinae. Likewise, the tribe
46Theroniini is resurrected for the *Theronia*-group of genera (stat. rev.). Phylogenetic analyses,
47in which we gradually increased the numbers of genes, revealed that the initially steep
48increase in mean clade support slows down at around 40 genes. Consideration of up to 93
49genes still left various nodes in the inferred phylogenetic tree poorly resolved. It remains to be
50shown whether more extensive gene or taxon sampling can resolve the early evolution of the
51pimpliform subfamilies. However, the fact that the single-gene analyses did not reveal strong
52incongruences, but instead all showed low resolution at the base of Pimpliformes, suggests
53the possibility of a hard polytomy.

54

55Keywords: next-generation sequencing, hybrid capture, phylogenomics, phylogenetic
56informativeness, classification

57Introduction

58The Ichneumonidae constitute the largest family of parasitoid wasps, with to date more than
5925,000 species described and conservative estimates suggesting that several fold more are yet
60to be discovered . Ichneumonid wasps in the vast majority of cases attack immature stages of
61holometabolous insects and, to a lesser extent, immature and mature spiders . They can also
62be predators of arachnid egg sacs and even more rarely, facultatively herbivorous . They play
63a crucial role in most terrestrial ecosystems by regulating the abundance of other species,
64including pests in agriculture and forestry . Despite the fact that there are probably many more
65ichneumonid wasps than vertebrates (about 68,600 species; the group has received only
66limited attention from systematists and evolutionary biologists and the phylogenetic
67relationships among and circumscription of higher-level taxa within the family are still very
68poorly understood . These limitations hinder inferences pertaining to evolutionary pathways
69of parasitoid life styles, major changes in host ranges, and co-evolution with their hosts. The
70current classification of the family is based mostly on morphological data , and the only
71available molecular phylogeny of the family is based on a single gene (28S rRNA) that is
72highly sensitive to alignment parameters . The current consensus in ichneumonid taxonomy
73and phylogeny was summarized by Quicke and Broad *et al.* . The authors pointed out that the
74current classification recognising up to 42 extant subfamilies will likely be subject to future
75changes. Seven informal subfamily groups are currently recognised, mostly based on larval
76morphology and on results from a phylogeny obtained from the combined analysis of the 28S
77rRNA gene and a morphological character matrix . Three of these seven groups contain most
78of the ichneumonid species: Ichneumoniformes, Ophioniformes, and Pimpliformes .

79 We here focus on the clade Pimpliformes, which includes nine subfamilies, the
80Acaenitinae, Collyriinae, Cyloceriinae, Diacritinae, Diplazontinae, Orthocentrinae,
81Pimplinae, Poemeniinae, and Rhyssinae. They show a great variety in their exploited host
82groups and life history strategies. Specifically, they include parasitoids of, in descending
83order of importance, Lepidoptera, Hymenoptera, Coleoptera, Diptera, and Araneae; they act
84as ectoparasitoids and endoparasitoids, and as idiobionts and koinobionts . Idiobionts stop the
85development of their hosts at the time of parasitisation, with the wasp female usually
86paralysing the host larva or pupa permanently before laying her egg on top of or, more rarely,
87inside it . Koinobionts, in contrast, allow their hosts to continue their development, often over
88several moulting cycles; they are typically endoparasitoids, although the subfamily
89Tryphoninae and the *Polysphincta* genus-group in Pimplinae are conspicuous exceptions in
90being koinobiont ectoparasitoids . All four strategies are represented by several subfamilies in

91the Pimpliformes, and several transitions have probably taken place, both between idiobiosis
92and koinobiosis and between ecto- and endoparasitism within this group. Also, a single
93transition to Diptera as a host group has been postulated , but the phylogenetic relationships
94among the pimpliform subfamilies are too uncertain to draw any robust conclusions .

95 In addition to studying subfamily relationships within Pimpliformes, we also use this
96study to examine operational questions regarding molecular phylogenetic analyses. Namely,
97we ask: how many and which genes are needed to resolve the pimpliform tree? A much-cited
98study from the early days of phylogenomics examined the relationships among eight species
99of yeast as inferred from 106 genes (comprising 127 kilobases (kb)) and concluded that "very
100large concatenated data sets of ~20 genes were required to provide 95% bootstrap support for
101all nodes" . This gene number might seem low now, given that hundreds to thousands of
102genes are regularly sequenced to infer large phylogenies . However, the discussion triggered
103by this study focused on two questions that are now more timely than ever: what are the
104causes of systematic biases in phylogenomics, and how many genes are really needed if they
105are chosen carefully?

106 The importance of the above questions is underlined by the increasing number of
107phylogenomic studies that arrive at contradictory conclusions, despite including hundreds to
108thousands of genes . The identification of systematic biases often requires re-analysis of the
109entire dataset, which can be prohibitive for very large datasets. The same applies to analyses
110that aim to study the influence of different taxon sampling schemes to assess the robustness of
111phylogenetic inference . Thus, increasing dataset size often comes at the cost of reduced
112rigour in the analysis methodology. The computational limitations imposed by many of the
113more advanced analytical methods, such as certain Bayesian inference methods, and the use
114of complex evolutionary models, represent another important argument for carefully choosing
115markers instead of simply increasing dataset size.

116 Several criteria have been suggested to aid in identifying the phylogenetically most
117informative genes, such as i) previous performance in a different, typically smaller taxon set ;
118ii) evolutionary rate and its distribution among sites ; iii) clock-like evolution ; and iv)
119stationarity of nucleotide or amino acid composition . However, different studies of empirical
120datasets have arrived at conflicting conclusions about the performances of these criteria , with
121the consequence that the pathway to efficient experimental design in phylogenomics remains
122obscure. It is even likely that identification of the best-performing approach will remain
123highly context-specific, with small differences in taxon sampling and analytical approach

making a large impact on the preferred strategy. Thus, only few criteria might emerge that have general applicability.

In any case, increasing the number of genes sampled can be expected to have a positive effect on the robustness of phylogenetic inference, at least up to a point, for instance by allowing identification of regions of uncertainty in a phylogenetic hypothesis that are due to incongruences among the individual gene trees. Next-generation sequencing facilitates the acquisition of unprecedented amounts of data, and techniques such as hybrid capture (also called anchored/hybrid enrichment or target DNA enrichment) enable cost-effective application even to non-model taxa. Hybrid capture approaches typically use RNA baits of about 100–150 bp in length, which show a close match to the target DNA sequences, in order to enrich the target DNA in a genomic extract. The enrichment step facilitates the pooling of large numbers of samples on a single sequencing lane, thus decreasing sequencing costs considerably. The technique has been applied to taxonomic groups with various levels of divergence, with capture success often still high in distantly related groups. Given that only a handful of markers have been established for ichneumonid phylogenetics in the past, and divergences within the Ichneumonidae likely date back to the upper Cretaceous, a hybrid capture approach is very promising for robustly inferring the phylogeny of this family.

We here aimed to resolve the higher-level relationships within Pimpliformes to gain insights into the evolution of this group, especially with respect to parasitoid strategy and host range. To this end, we used previously and newly generated transcriptome data to identify candidate genes, filtered them by several criteria of phylogenetic informativeness, and created a set of baits to enrich 152 exons of a total of 93 genes. We used this bait set to enrich the target DNA in 55 species of ichneumonids and two distantly-related outgroup taxa. Furthermore, we designed and tested oligonucleotide PCR primers for amplifying and subsequently Sanger-sequencing eight relatively long exons, identified in the captured gene set, which can be used in future ichneumonid studies. We applied several methods for phylogenetic reconstruction and examined them for congruence and conflict. Furthermore, based both on random and non-random gene sampling strategies, we studied subsets of our data to assess strategies to improve phylogenetic studies of the group in the future.

Materials and Methods

This study included five major steps: i) transcriptome sequencing and analysis, ii) estimation of phylogenetic informativeness of genes, iii) hybrid capture laboratory work and

156data analysis, iv) development and testing of oligonucleotide primers for PCR and Sanger
157sequencing, and v) phylogenetic analyses. These individual steps are detailed below.

158 Most of the bioinformatics for this study were conducted using newly developed
159scripts, which are provided here as supplementary material (Supplementary File S1). We
160mostly used shell scripting (bash) to create a pipeline of programs for specific tasks and the
161APE package in the program R to manipulate alignments and trees. All calculations were
162performed on UBELIX (<http://www.id.unibe.ch/hpc>), the HPC cluster at the University of
163Bern, Switzerland.

164*Transcriptome data, orthology prediction and transcriptome phylogenies*

165 Transcriptomes of two pimpliform wasp species were sequenced and published by
166Peters *et al.* . We additionally sequenced whole body transcriptomes of adult specimens for
167eight additional species from five pimpliform subfamilies and three other ichneumonids,
168including *Xorides praecatorius* Fabricius as a representative of the Xoridinae, that has been
169hypothesized to be sister to all remaining ichneumonids . For a full list of specimens used in
170this study, see Table 1 (and the more detailed Supplementary File S2). As we sampled a
171single species per genus in most cases, we henceforth refer to genera only. The specimens for
172the transcriptome analysis were freshly collected in a mountainous area in the canton of Bern,
173Switzerland, and either identified alive after a cooling period in the fridge, or collected
174alongside other, identical-looking individuals at the same location, with the latter kept in
175ethanol as ‘pseudo-vouchers’ or ‘paragenophores’ for later confirmation of the preliminary
176identifications (Supplementary File S2). Specimens chosen for the transcriptome analysis
177were either ground directly in RNAlater solution or shock-frozen at -80 °C. Total RNA was
178extracted using the Promega ReliaPrep™ Tissue Miniprep system. Library preparation was
179done using the Illumina TruSeq stranded RNA method, which includes poly-A selection,
180followed by size selection. The resulting 11 libraries were indexed, multiplexed and
181sequenced on one Illumina HiSeq3000 lane (150 bp paired-end, 250 million reads) at the
182Institute of Genetics of the University of Bern (VetSuisse).

183 The raw reads from transcriptome sequencing were de-multiplexed and trimmed using
184Trimmomatic 0.33 , which removed adapters, leading and trailing bases below a phred score
185of 10, and sections in a sliding window of size 4 with an average score below 15. Read quality
186was assessed with FastQC 0.11.2 . Transcriptomes were assembled with Trinity 2.1.1 in
187paired-end mode on six CPUs and with a maximum memory setting of 24 Gb of RAM under

188 default settings. Combined with transcriptomes from the 1KITE project , this resulted in 19
189 ichneumonid and six braconid transcriptomes (Table 1). The raw reads of the newly
190 sequenced taxa are archived at the National Center for Biotechnology Information (Genbank),
191 NIH, under BioProject PRJNA450386.

192 To identify orthologous transcripts of putative single-copy protein-coding genes
193 among these 25 transcript libraries, we searched them for 3,260 genes that the OrthoDB
194 database suggested to be single-copy in Hymenoptera . For this purpose, we used the
195 software Orthograph , which employs the best reciprocal hit criterion, and the ortholog set
196 designed by Peters *et al.* . Orthograph was run with the same settings and the identified
197 transcripts were post-processed in the same manner (*e.g.*, stop codons were masked with
198 NNN) as outlined by Peters *et al.* , except that we removed the DNA sequences of the
199 following species, which were part of the ortholog set: *Acromymrex echinator* (Forel), *Apis*
200 *mellifera* Linnaeus, *Camponotus floridanus* (Buckley), *Harpegnathos saltator* Jerdon, and
201 *Tribolium castaneum* (Herbst). We retained the DNA sequences of the jewel wasp *Nasonia*
202 *vitripennis* (Walker), whose DNA and amino acid sequences were also part of the ortholog
203 set. We exploited the gene modes of this species to cut the aligned DNA sequences of
204 orthologous transcripts at putative exon-exon boundaries during target DNA enrichment bait
205 design (see below). All orthologous DNA sequences were aligned at the amino acid level
206 (provided by Orthograph) with MAFFT version 7.123 applying the L-INS-i alignment
207 algorithm. The alignments obtained were used as blueprints to align the corresponding
208 nucleotide sequences with a modified version of Pal2Nal version 14.1 .

209 We used a custom R script to trim leading and trailing amino acid positions in the
210 resulting multiple DNA sequence alignments of each gene to ensure that the flanking
211 positions had at least 90% coverage across the 26 included taxa. To obtain candidate loci for
212 which to assess phylogenetic informativeness, we used a coverage filter, retaining only genes
213 with at least 600 bp in total, at least 300 bp present in every taxon, and at most 5% of the
214 alignment consisting of gaps or missing data. Only 723 genes fulfilled the above criteria and
215 their alignments are provided as supplementary material (Supplementary File S3). In order to
216 ensure that there was no cross-contamination between taxa resulting from the pooled Illumina
217 sequencing procedure, we checked the alignments of the 723 genes for raw pairwise distances
218 below 1 % and found only a single case which can be explained by a true close relationship
219 (0.87 % between *Dolichomitus imperator* and *D. quercicolus* in gene Nasvi2EG0030079). To
220 examine the phylogenetic signal in the transcriptome data, we conducted phylogenetic

analyses on this gene set with RAxML 8.1.2 , both at the amino-acid and nucleotide levels. Partitioning schemes and amino acid substitution models for each partition were obtained using PartitionFinder2 with the rcluster algorithm using equal weights on the clustering criteria rate, state frequencies, model parameters, and among-site rate variation. Since RAxML only allows for a single model across partitions when analysing nucleotide sequence data, even though the model parameters can be estimated separately for each partition, we applied the GTRCAT model when analysing the nucleotide alignments. When analysing the amino acid sequences, we applied the identified fixed-rate model from PartitionFinder in combination with the CAT model of among-site rate variation. To assess strength in signal, we ran 100 non-parametric bootstrap replicates. As all nodes obtained maximum bootstrap support, we also conducted gene jackknifing by randomly sampling sets of 10 genes 100 times. Each of these datasets was analysed in RAxML as amino acid alignments (under the PROTCATJTT substitution model, the most commonly selected model by PartitionFinder on the full dataset) and as first plus second codon positions (using the GTRCAT nucleotide substitution model), partitioning by gene and/or codon position.

Estimating phylogenetic informativeness

To obtain predictors of phylogenetic informativeness for the 723 candidate genes, we first trimmed each of them down to 600 bp that showed the best taxon coverage. We then obtained five predictors: achieved mean clade support per node, ML quartet mapping score, sites with optimal evolutionary rates, clocklikeness, and the stationarity of nucleotide frequencies. First, to get a measure of clade support, we ran single-gene ML analyses under the GTR + Γ + I model , partitioning by codon position and with 50 bootstrap replicates. We then recorded the sum of bootstrap support values for 12 nodes which were unequivocally retrieved in all phylogenetic analyses of the concatenated transcriptome data and that are in agreement with taxonomy and with previous phylogenetic studies : Braconidae, Ichneumonidae, Ichneumonidae without *Xorides* Latreille, Ophioniformes, Ichneumoniformes, Pimpliformes, (Pimplinae + Rhyssinae + Poemeniinae), Pimplinae, (Rhyssinae + Poemeniinae), Pimplini, and the genera *Pimpla* Fabricius and *Dolichomitus* Smith. We refer to this measure as the ‘taxonomic bootscore’. We also separately recorded if the rooting of the Ichneumonidae was in accordance with a previous molecular and morphological analyses , i.e., whether *Xorides* was recovered as the sister group to all other ichneumonids.

253 Maximum likelihood quartet mapping has been shown to perform well as a predictor
254 of phylogenetic informativeness in datasets simulated at varying evolutionary rates . We
255 obtained treelikeness scores under the HKY model for 1,000 randomly chosen taxon quartets,
256 using a custom R script (Supplementary File S1).

257 To estimate evolutionary rates at each site of the alignment, we first transformed the
258 tree obtained from the ML analysis of the complete amino acid dataset into an ultrametric tree
259 using the penalized likelihood function 'chronopl' from the APE package in R . An ML
260 estimate of the transition/transversion ratio was obtained for each gene using this tree and
261 APE's 'optim.pml' function. Transition/transversion ratios and an ultrametric tree were both
262 used as input for the program DNARates , which calculates site-specific evolutionary rates
263 under ML. The number of sites which evolve at a rate between 0.1 and 0.4 expected
264 substitutions per root-tip distance was recorded for each gene, a range found to result in near-
265 optimal performance in a simulation study .

266 Clocklikeness of the genes was measured by running single-gene relaxed-clock
267 analyses under the independent gamma rates (IGR) model in MrBayes 3.2 . The data were
268 partitioned into first plus second versus third codon positions and analysed under the GTR + Γ
269 + I model, with the topology constrained to include the taxonomic groups that comprised the
270 taxonomic bootscore (see above) to avoid convergence issues and obtain comparable results.
271 The median estimate of the variance of the clock model ('igrvar' parameter) was used as a
272 measure of clocklikeness.

273 To test for deviations from the stationarity hypothesis of nucleotide composition, we
274 used the Chi-square test of homogeneity as implemented in PAUP* 4.0 . This test does not
275 account for phylogenetic signal in the data, but we assume that it is conservative and thus
276 sufficient for our purpose of identifying and excluding genes that show a large variation in
277 nucleotide composition across taxa.

278 Aiming for a good balance between the number of enriched genes and pooled taxa, we
279 developed baits of 120 bp length each for 100 candidate genes using the program BaitFisher .
280 In order to span the sequence variation that can be expected in such a taxonomically broad
281 sample, we included separate baits for each cluster of sequences with at least 0.12 nucleotide
282 sequence divergence for a specific bait region. We used the intron-exon boundaries from the
283 *Nasonia* genome to inform bait positions and, in a second step, filtered the obtained baits
284 against the genome of *Microplitis demolitor* Wilkinson , a member of the Braconidae and, at

285least at the time of our bait design analyses, the closest relative to our ingroup with an
286annotated genome. Baits were only retained if they had a single hit and 95% overlap in the
287*Microplitis* genome. We applied different tiling designs: we tiled five baits with an offset of
28820 bp across a target coding exon in 51 genes (68 exons), three baits with an offset of 20 bp
289across a target coding exon in 36 genes (77 exons), and a single bait to capture a given target
290coding exon in six additional genes (8 exons). In total, we designed 5,419 baits to capture 153
291exons for a total of 93 genes. Bait nucleotide sequences are provided in FASTA format in the
292supplementary material (Supplementary File S4) and were ordered as RNA baits from the
293company MYcroarray (now Arbor Biosciences, Ann Arbor, USA).

294*Hybrid enrichment laboratory work*

295 The wet laboratory procedure was as follows: genomic DNA was extracted using the
296Gentra Pure-gene extraction kit (Gentra systems, Minneapolis, MN USA). Extractions were
297from either single legs or whole bodies, with the extracted specimens kept as vouchers in a
298number of museum collections (Supplementary File S2). DNA concentration was measured
299using fluorometric quantification (Invitrogen Qubit) for all samples and for about a quarter of
300the samples also with the Agilent 2200 TapeStation System. DNA was sheared to an average
301fragment distribution of about 400 bp using a Diagenode Bioruptor sonicator, with the
302shearing success subsequently verified on the TapeStation. Libraries were constructed
303following Meyer and Kircher , using reagents from ThermoScientific and unique
304combinations of dual i7 and i5 index sets to allow for subsequent pooling. BeckmanCoulter
305AMPure XP beads were used for all purifications following the bead ratios specified by
306Meyer and Kircher . We assessed the success of library preparation by measuring DNA
307concentration with fluorometric quantification and for about a third of the samples also with a
308standard qPCR run using the LightCycler 96 Real-Time PCR System. In some of the libraries,
309we detected adapter dimers after running them on a standard electrophoresis gel. We cut these
310from an agarose gel, which had been stained with GelGreen, and purified the gel slices
311excluding the adapter dimers using Qiagen MinElute Gel Extraction columns and the
312manufacturer's protocol.

313 Prior to capture, we divided the target baits into $\frac{1}{2}$ capture reactions. We enriched one
314to four samples per $\frac{1}{2}$ capture depending on importance of the taxa and similarity of the DNA
315concentration in the libraries. The target capture was performed with a 24 h hybridization at
31665 °C following the MYbaits protocol v3 from Microarray. The post-capture amplification
317was completed using KAPA HiFi DNA Polymerase (Kapa Biosystems, USA) with the

318following PCR protocol: 98 °C for 30 s, followed by 12 cycles of 98 °C for 20 s, 60 °C for 30
319s, and 72 °C for 1 min, with a final extension of 5 min at 72 °C. Illumina IS5 and IS6 primers
320were used to attach to the i7 and i5 ends of the library.

321 We verified enrichment success with a final run on the Agilent 2200 TapeStation
322System. Size selection was performed using AMPure XP beads in order to remove the small
323nucleotide fragments that may have been introduced during the capture. Samples were
324multiplexed into three pools of 14, 24, and 23 samples at equal concentrations and were run
325on one lane each on the Illumina MiSeq (pool 1) and the Illumina Next-Seq platforms (pools
3262 and 3), obtaining 150 bp paired-end reads (Australian Genome Resource Facility (AGRF) in
327Adelaide, Australia).

328*From raw reads to alignments*

329 We closely followed the strategy described by Bragg *et al.* , which is tailored for
330hybrid capture data from highly divergent taxa, but developed our own scripts. All scripts
331used in our bioinformatics pipeline are publicly available, including detailed comments
332(Supplementary file S1).

333 The paired-end reads were de-multiplexed and partial adapter sequences removed
334using the program Cutadapt . We used Trimmomatic to remove leading and trailing bases
335below a phred score of 30 and sections in a sliding window of size 4 with an average score
336below 25. Read quality was ascertained with FastQC 0.11.2 . BlastX was used to search for
337reads in each sample that showed a significant match with each of the enriched genes, a
338strategy making use of the more conserved amino-acid regions in comparison to nucleotide
339sequences across taxa separated by long divergence times . Only after this step were reads
340assembled into contigs separately per gene and sample, first, by using the program Velvet to
341generate contigs under a range of Kmer values (25 to 75 bp), and second, by merging (some
342of) the resulting contigs with the software Cap3 . To ascertain that the identified contigs
343indeed represent the targeted single-copy genes, we used BlastX to align the contigs to the
344predicted mRNA of the *Nasonia* official gene set , only retaining best matches. For most
345genes, this resulted in more than a single contig, as introns were often too long to be fully
346captured. To identify intron-exon boundaries, we used the 'protein2genome' method in the
347aligner Exonerate with all the transcriptome sequences as queries, keeping the longest match.
348Exonerate outputs the sequence identities of its matches, and we removed those that were
349below 60%, a threshold obtained by comparing the transcriptome sequences among each

other. After translation into amino acids, the remaining matching contigs were aligned to all transcriptomes and to each other in the multiple-sequence aligner MAFFT using the E-INS-i algorithm. The contigs were then translated back into nucleotide sequences using a custom R script. As Exonerate only finds complete introns, and contigs often started and/or ended in intronic sections, we developed an R script that identifies leading and trailing intron portions based on a drop in the pairwise distance when intron-exon boundaries are crossed. This script identified many but not all remaining intronic sequences, leaving out mostly those overhangs which were too short to calculate a meaningful distance. Thus, we had to correct some alignments manually (a log of these changes is given as Supplementary File S5). The remaining contigs were used as queries in a BlastN search against the original reads with a restrictive e-value of 10^{-10} to obtain coverage values and perform variant calling. Sequences were retained from a minimum coverage of eight identical bases, with variants called as ambiguity codes at a minimum coverage of four or at least 10% of the total coverage, with most positions obtaining much higher coverage (Supplementary File S6). The alignments of all 93 genes are provided as supplementary material (Supplementary File S7).

Primer development for Sanger sequencing

One of the main hindrances for developing PCR oligonucleotide primers for amplification and direct Sanger sequencing of nuclear protein-coding genes is the short average length of exons in most arthropod genes; this is problematic because both the length variability and high AT content of introns can pose problems for successful PCR and sequencing. We initially attempted to develop primer pairs for the eight genes with the highest phylogenetic informativeness, but in almost all cases found several introns ranging up to more than 1,000 base pairs in length within each of the target segments. To overcome this issue, we used the intron-exon boundaries identified in the *Nasonia* genome, together with the data obtained from the hybrid enrichment, to search for long exons (> 300 bp) or for exons separated by an intron that was relatively short (less than about 200 bp in most of the taxa whose genes we captured via hybrid enrichment; Table 2).

Primer design was done manually in Geneious 7.1.3 based on the *Nasonia* genome and the obtained ichneumonid transcriptome and hybrid capture sequences, while primer synthesis was by Microsynth (Balgach, Switzerland). The primer pairs (Table 2) were tested in a three-step procedure: i) touchdown PCR with a provisional annealing temperature on eight taxa sampled across ichneumonids; ii) PCR without touchdown with a temperature gradient on four new taxa; iii) touchdown PCR at the optimal temperature from step 2 on five new and

three previously failed taxa. The PCR mixtures (25 µl) contained 10 pmol of each forward and reverse primer, 12.5 µl GoTaq® Hot Start Green Master Mix (Promega, USA) and 1.5 µl DNA extract. PCR protocols consisted of an initial denaturation for 3 min at 94 °C, followed by 36–37 cycles of 30 s at 94 °C, 45 s at optimal annealing temperature (Table 2), 1–1.5 min at 72 °C, followed by a final extension step for 10 min at 72 °C. For the touchdown protocol, two cycles with annealing temperature 5 °C higher and two cycles with annealing temperature 2 °C higher than optimal were added before 36–37 cycles at annealing temperature. The successfully amplified DNA fragments were sequenced by LGC Genomics (Berlin, Germany). Finally, we applied the primers to a broad set of more than 100 ichneumonid and two non-ichneumonid species to ascertain broad applicability.

Phylogenetic analyses

The final set of alignments covering the 93 enriched genes was subjected to various phylogenetic analyses. First, we analysed it under the maximum likelihood (ML) criterion in RAxML 8.1.2 after determining partitioning schemes in PartitionFinder2, as described for the phylogenetic analyses of the transcriptomes. As RAxML only allows for a single substitution model for all nucleotide partitions, we ran the GTRCAT model for the nucleotide alignments, while the amino acid alignments were analysed under the partition-specific substitution chosen by PartitionFinder2 and the CAT model of among-site rate variation. To assess signal strength, we ran 1,000 non-parametric bootstrap replicates. The concatenated genes were analysed at the nucleotide level by including i) all codon positions and ii) first and second codon positions only. We additionally analysed the concatenated dataset at the amino acid level. The amino acid alignment was also subjected to Bayesian analyses using fixed amino-acid substitution models (as obtained from PartitionFinder2) in MrBayes 3.2 and the empirical profile mixture model implemented in PhyloBayes-MPI 1.7. In the analyses with MrBayes, we increased the relative probability of the topology and branch-length moves to 200 in order to counter-balance the large number of substitution model parameters estimated separately for the 56 partitions. In both Bayesian analyses, we conducted four independent runs to assess convergence.

Reconstructing the evolution of parasitoid lifestyle

As an example of an important aspect of the parasitoid lifestyle, we used ancestral state reconstruction to examine transitions between idiobiosis and koinobiosis in

415ichneumonids. Biological data for the included genera were extracted from several sources .
416For many ichneumonid genera, no sufficiently detailed records of their parasitoid strategy
417exist, but if such information was available, it typically indicated that closely related genera
418and even members of the same subfamilies shared the same strategy. Thus, we extrapolated
419from related taxa in several cases (see Supplementary File S8 for details). Ancestral states
420were reconstructed under maximum parsimony using the ‘hsp_max_parsimony’ function in
421the R package ‘castor’ , which allows for reconstruction of missing tip data. We assessed the
422impact of topological uncertainty on the number and direction of evolutionary transitions by
423including consensus and ML trees from the various Bayesian and ML analyses, plus 1,000
424trees each from the Bayesian posterior distributions as obtained from the analyses of the
425amino acid alignments with PhyloBayes and MrBayes, respectively.

426*Increase of tree resolution with dataset size*

427 To assess the relationship between the number of genes included and the mean clade
428support of the resulting tree, we randomly sampled two to 90 genes, with 20 replicates for up
429to 20 genes and 10 replicates from 30 genes. Each concatenated gene set was analysed in
430RAxML at the nucleotide level, partitioned by codon position and using the GTRCAT model
431with 50 bootstrap replicates. Mean clade support was measured as the sum of bootstrap
432support values above 50%.

433 In order to re-evaluate the success of the criteria ‘clade support’, ‘clocklikeness’ and
434‘evolutionary rates’, we estimated their scores again under the full taxon sampling for each
435gene separately. For the clade support measure, RAxML analyses were conducted at
436nucleotide level, partitioning by codon position and under a GTRCAT model. Genes were
437ranked according to decreasing mean clade support as obtained from 100 bootstrap replicates.
438To obtain a measure of the clocklikeness of the genes, we ran Bayesian relaxed-clock
439analyses under the independent-gamma-rates model (IGR) in MrBayes 3.2 . We used the
440uniform branch length and topology prior and an exponential distribution with rate 1 as the
441prior on the clock variance. As the topology signal in some single-gene datasets was quite
442low, we used three constraints in order to force a reasonable rooting in the clock analyses :
443Ichneumonidae, Braconidae, and Ichneumonidae + Braconidae. Analyses were run for 10 Mio
444generations, a value that was increased to 15 Mio generations if the average standard
445deviation of split frequencies (ASDSF), a measure for topology convergence, had not dropped
446below 0.02. The genes were then sorted according to descending median clock variance. As a
447third criterion, we estimated site-specific rates in the program DNARates as previously for

the transcriptome sequence data and ranked the genes according to the number of sites in the ‘near-optimal’ rates bracket .

Results

Transcriptomes and phylogenetic informativeness

An average of 10^8 bp ($8.2 \times 10^7 - 1.3 \times 10^8$) were sequenced from each transcriptome library, which assembled into ca. 10^5 contigs ($7.2 \times 10^4 - 1.5 \times 10^5$). Both these numbers are about 3.6 times larger than those of the transcriptome data published by Peters *et al.* (Supplementary File S9), which can be explained by the fact that we pooled fewer samples per lane. Searching for 3,260 single-copy genes, we found comparable numbers in our transcriptomes and those published by Peters *et al.* (3,059 compared to 2,969). However, different genes were missing in different taxa, so that only 1,908 genes were present across all 25 transcriptomes. Filtering these for sections of at least 600 bp lengths and good coverage across all taxa resulted in 723 genes spanning more than 568 kbp in length.

Analysing the above 723 genes both as concatenated amino acids, first and second codon positions, and full nucleotide alignments under ML, we recovered partly conflicting trees (Fig. 1), each of which obtained maximal bootstrap support at every node. Such a result might be expected from a taxon sampling as sparse as this one, even when a large number of genes are analysed, and thus we focus on the few well-supported relationships and on identifying the main regions of conflict among the transcriptome trees. As expected from previous studies, *Xorides* was recovered as the sister to the remaining ichneumonids and the higher taxonomic groupings Ophioniformes, Ichneumoniformes, and Pimpliformes were recovered in all three datasets. However, the relationships among the pimpliform subfamilies showed conflicts and the subfamily Pimplinae, which was recovered as monophyletic in the amino-acid analysis, was split into two groups in the nucleotide analyses, both when including and when excluding third codon positions. The two analyses at the nucleotide level showed a single conflict in regard to the placement of *Rhyssa* Gravenhorst either as sister to *Deuteroxorides* Viereck + Pimplinae II (Fig. 1) or as sister to (*Collyria* Schöidte + *Coleocentrus* Gravenhorst) + *Dolichomitus*. In order to identify regions in the tree with lower support, we performed gene-jackknifing based on 100 datasets of 10 randomly chosen genes each, analysed at the amino acid level or including first and second codon positions. These analyses recovered mostly, but not always, the same nodes that were in conflict between datasets as less well supported in the respective trees. These nodes also tended to be separated by comparatively short branches in all analyses (Fig. 1).

481 The 723 candidate genes were filtered based on their degree of phylogenetic
482informativeness, considering five measures (Table 3): i) no significant deviation from
483compositional stationarity in the nucleotides, including third-codon positions (p-value of Chi-
484square test above 0.05); ii) clock variance below 0.015 ; iii) root of Ichneumonidae correct
485(*Xorides* recovered as sister to the remaining ichneumonid taxa); and (iv) at least 80 of the
486600 nucleotides evolving at a near-optimal rate. This filtering approach left 117 genes, which
487were ranked by the taxonomic bootscore (sum of bootstrap support for 12 chosen groups
488recovered in the 723-gene and previous analyses, see Materials and Methods), and the 100
489best genes were chosen for the hybrid capture approach. Seven of these turned out not to
490include any exons that were long enough for bait design, at least not among the exons for
491which we had sufficient taxon coverage, which left a set of baits to capture 152 exons for a
492total of 93 genes.

493Hybrid capture success and Sanger sequencing primers

494 All 93 target genes were enriched successfully, with an average length of the captured
495coding target sequence of 376 bp per taxon after quality and coverage filtering (across genes:
496median = 415 bp, min = 150 bp, max = 1,197 bp; see Supplementary File S9). Calculated
497across all 59 taxa in the hybrid capture approach, the minimum coverage across the target
498genes was 331 bp, and on average 90 genes were recovered with more than 100 bp. The
499exception was the outgroup sample PTE_cf_*Eupelmophotismus* from the superfamily
500Chalcidoidea, which had much lower enrichment success (108 bp on average across genes,
501and only 19 genes with more than 100 bp). It is unclear if this was due to its taxonomically
502more distant position from the taxa on whose DNA the bait design was based on, or due to the
503low quality of the DNA extract from the rather small specimen that we studied. Another even
504more distantly related outgroup, *Gasteruption* Latreille (Evanoidea), showed only slightly
505lower success when compared to the ingroup samples (337 bp on average per gene, 83 genes
506with more than 100 bp). Comparing the coverage of the hybrid capture with the transcriptome
507sequence data, we found that the latter performed better, with an average of 436 bp across
508taxa and genes. The full alignment of 93 genes for 84 taxa contained 41,565 bp with a gap
509proportion of 14%, a value that dropped to 12.4% when only the 75 ichneumonid taxa were
510included.

511 Trying to identify comparatively long exons, in the entire 93-gene dataset we found
512only four exons that were longer than 450 bp and eight more that were longer than 350 bp.
513Nevertheless, we developed oligonucleotide primers for eight genes that amplified and

514sequenced with good consistency (Table 2, Supplementary File S10). In one case (gene
515NasViEG013087), a combination of two primer pairs were needed to obtain target amplicons
516from DNA of the majority of taxa. In two cases, the amplified fragments spanned one intron,
517which was almost consistently about 100 bp in length, even in distantly related taxa. The
518exonic portion of the fragments ranged between 328 bp and 536 bp (Table 2) and was
519amplified in species across the entire ichneumonid tree and represented most of the tested
520subfamilies. Success was lower in the two specimens from different families (Braconidae and
521Gasteruptiidae), but additional testing is needed to see if that was due to low quality of the
522extracts or divergent sequences.

523Phylogeny of Pimpliformes

524 The five different phylogenetic analyses that we performed recovered different trees
525depending on data type and analysis method, with most incongruence located close to the base
526of the pimpliform radiation (Figs 2 and 3). This area of the resultant phylogenies also
527exhibited very short branches. The incongruent nodes were mostly the same as those
528conflicting between the nucleotide versus amino acid versions of the transcriptome analyses,
529but appeared much more complex because of the inclusion of additional subfamilies in the
530hybrid capture taxon sampling (Fig. 1). For several nodes at the base of the pimpliform
531radiation, no analysis retrieved a significant signal (clade support below 0.95 posterior
532probability or 85% bootstrap support: Fig. 3), and most of the gene trees were also unresolved
533at these nodes (Supplementary File S11). In contrast, the base of the inferred tree was well
534resolved. *Xorides* as the representative of Xoridinae consistently appeared at the base of a
535monophyletic Ichneumonidae, and the three informal subfamily groups Ophioniformes,
536Ichneumoniformes and Pimpliformes were recovered with maximum support in all analyses.
537The only exception from this consistently well-supported resolution was the placement of the
538only representative of the Labeninae (*Poecilocryptus* Cameron), which branched off either
539right after the branch leading to *Xorides* (Figs 3b, c), Ophioniformes (Figs 3a, d, e), or
540Ichneumoniformes (Fig. 3f). The latter would imply a sister group relationship between
541Labeninae and Pimpliformes.

542 The monophyly of most pimpliform subfamilies was well-supported by all analyses.
543Exceptions were the Pimplinae (because the genus *Xanthopimpla* Saussure and sometimes
544also the tribe Pimplini took up a position closer to the base of the pimpliform radiation), and
545Acaenitinae (with the tribe Acaenitini recovered separately from *Coleocentrus* in the tribe
546Coleocentrini). Pimplinae (excluding *Xanthopimpla*) retrieved maximum support only in the

547amino acid analyses (Figs 3a, b, c, d), while Orthocentrinae were only supported by Bayesian
548inference analyses (Fig. 3a, b, c). The only relationships among higher groups of
549Pimpliformes that were recovered consistently were sister group relationships between
550Rhyssinae and Poemeniinae, between the pimpline tribe Delomeristini and the *Theronia*
551genus-group, and (in all but the ML analysis of codon positions 1 and 2) between the
552Orthocentrinae and Diacritinae. The pimpline tribe Ephialtini was consistently recovered as
553monophyletic, which was also true for Pimplini (excluding the *Theronia* genus-group) and for
554Delomeristini (including *Pseudorhyssa* Merrill).

555 Three taxa were consistently recovered in places that are in conflict with their current
556classification (Figs 2 and 3). The genus *Pseudorhyssa*, regarded as a member of Poemeniinae
557since Gauld (1991), was placed with maximum support in all analyses in the tribe
558Delomeristini (Pimplinae). The *Theronia*-group of genera (represented in our analyses by one
559species each of *Theronia* Holmgren, *Nomosphesia* Gupta and *Parema* Gupta) appeared as the
560sister to the Delomeristini, including *Pseudorhyssa*, while *Hemiphanes* Förster, which has
561been classified as a member of the informal *Helictes* genus-group within Orthocentrinae,
562grouped with the Ichneumoniformes, a placement that also received maximum support, even
563though its position within Ichneumoniformes is not resolved. In the cases of both *Hemiphanes*
564and *Pseudorhyssa*, the new placements were also recovered in most of the single-gene trees,
565while they were never placed with the subfamilies they are currently classified in
566(Supplementary File S11). The placement of *Pseudorhyssa* with Delomeristini was supported
567in 89 out of 93 gene-trees, 44 of them with bootstrap support of at least 85%, and that of
568*Hemiphanes* with Ichneumoniformes by 81 gene-trees, 22 of them with at least 85% bootstrap
569support. The *Theronia*-group showed more variation in its placement among the single-gene
570trees, appearing with Delomeristini in 25 and with Pimplini in 15, while it showed various
571placements in the remaining gene trees (Supplementary File S11).

572Inferring the evolution of parasitoid life styles

573 Despite the low resolution at the base of the pimpliform radiation, the maximum
574parsimony reconstructions of parasitoid ecology retrieved very similar numbers of transitions
575across analyses (Fig. 3, Table 4). However, the directions of change within Pimpliformes
576varied strongly among analyses, with the most recent common ancestor of Pimpliformes
577recovered as an idiobiont with high probability in the PhyloBayes and nucleotide analyses,
578while it was retrieved as a koinobiont in most trees for the MrBayes analysis of the amino
579acid data (Table 4; the bootstrap trees from the ML analysis at the amino acid level were

580equivocal). Nevertheless, it is interesting to note that the idiobiont and koinobiont
581Pimpliformes, tended to cluster together in most of the analyses, with the exception of
582*Xanthopimpla* (in most trees) and the derived *Polysphincta*-group of genera. For two included
583taxa, the mode of parasitoidism is unknown; the reconstruction for Diacritinae was always in
584favour of it being a koinobiont, while it was sometimes equivocal for *Coleocentrus*.

585*How many genes do we need?*

586 Random subsampling of the full dataset of 93 genes resulted in a curve of increasing
587mean clade support with increasing dataset size (Fig. 4). The curve first showed a very steep
588increase up to about 40 genes, after which the improvement slowed down. Even at 93 genes,
589only a 90% mean clade support was achieved, with lower values mostly due to the basal
590pimpliform nodes (Figs 2 and 3). However, it is unclear whether saturation was reached,
591which suggests that additional gene sampling could lead to a further increase in tree resolution
592even under the same taxon sampling and analytical method.

593 When not sampling genes randomly but instead ranking them according to different
594criteria, we found a clear improvement for up to about 30 genes under all criteria except the
595ranking by clocklikeness (Fig. 4). The rankings of the different criteria were very highly
596correlated (Supplementary File S12), with the highest correlation coefficients between gene
597length and the number of sites evolving at a near-optimal rate, while clocklikeness showed the
598weakest correlation with the other criteria. Thus, gene length seems to be the main factor
599determining differences in single-gene success in our hybrid capture dataset, even though the
600evolutionary-rate and clade support criteria performed slightly better between 5 and 10–30
601genes (Fig. 4).

602 When comparing the ranking obtained from the full hybrid-capture dataset with the
603gene rankings obtained from the transcriptome sequence data, we found no significant
604correlation between gene rankings, even though the bootscore and mean clade support
605measures were weakly correlated with each other (Spearman: $Rho = 0.18$, $p = 0.08$;
606Supplementary File S12). Ranking genes according to their performance in the transcriptome
607sequence data also did not lead to any improvement over random gene sampling (results not
608shown). Besides the obviously different taxon sampling, the two datasets also differed
609strongly in gene length, as the hybrid capture was only enriching some of the exons, while
610nearly complete genes were available in the transcriptome sequence dataset.

611**Discussion**

612 *New molecular resources for Ichneumonoidea*

613 The bait set for hybrid enrichment proposed here proved highly efficient in enriching
614 152 exons of 93 genes, not only across the target group, the ichneumonids, but also in the
615 sister family Braconidae and in the genus *Gasteruption*, a member of a distantly related group
616 of apocritan wasps (Evanioidea: Gasteruptiidae). Given the small number of genes previously
617 used in ichneumonid phylogenetics, this bait set represents a significant step forward. We
618 also provide oligonucleotide primers for PCR and Sanger sequencing of eight markers that
619 were tested successfully across Ichneumonidae and several outgroups. This set complements a
620 recent study very well which suggested oligonucleotide primers for apocritan Hymenoptera,
621 but did not include any representatives of Ichneumonoidea. The largest obstacle for
622 developing PCR primers is the large density of introns and corresponding short exon lengths
623 found in many arthropod taxa. Despite the fact that the divergence between *Nasonia* and
624 Ichneumonidae probably dates back to the Triassic, most of the intron-exon boundaries were
625 found to be conserved, which might be because introns typically insert at certain sequence
626 motifs. This high conservation of intron positions across groups encourages the use of
627 distantly related models for intron-exon boundaries for designing phylogenetic markers, be it
628 for PCR amplification or in bait design for hybrid captures.

629 Of the eight genes for whose amplification we developed PCR primers, two actually
630 included an intron, but one that was found to be short in almost all the taxa we sequenced.
631 Preliminary analyses showed that the captured intronic sequences could not be aligned at the
632 deep phylogenetic divergences found in our dataset, but they might prove useful in future
633 studies as markers for species delimitation. Given previous reports of shortcomings of the
634 standard mitochondrial barcode locus cytochrome oxidase subunit 1 (*COI*) in insects and
635 especially ichneumonids, additional nuclear markers that are sufficiently variable at the
636 species level, such as introns, are sorely needed. However, it remains to be shown how useful
637 our intronic markers are at delimiting species.

638 *Taxonomic implications of the pimpliform tree*

639 Our combined transcriptome and hybrid capture data represent an important step
640 forward from the previous molecular datasets that targeted higher ichneumonid relationships,
641 all of which included only a single gene (28S rRNA) and were highly sensitive to alignment
642 strategies. Even the addition of morphological data in Quicke *et al.* left many parts of the
643 ichneumonid tree inconsistent between different analytical approaches; this might be due to

high levels of homoplasy found in these parasitoid wasps, which can partly be explained by convergent adaptations to the same host groups .

Given the sparse taxon sampling in the transcriptome data, which omitted members of several pimpliform subfamilies, we focus our discussion here on the main analyses of the combined dataset, but mention results from the transcriptome trees where appropriate. Many previously suggested higher relationships, most of which were based on morphology, were confirmed here. This is case for the placement of Xoridinae , which is unequivocally recovered as sister to the remaining ichneumonids here for the first time . The informal subfamily groupings Ophioniformes, Ichneumoniformes and Pimpliformes, and the monophyly of most subfamilies, were also recovered with good support.

With respect to relationships among the Pimpliformes subfamilies, results of our amino-acid analyses (Figs 2, 3a-d) are similar to the "best-guess" phylogeny shown in Quicke ; which is a summary of the analyses in Quicke *et al.* , in terms of the grouping of Poemeniinae, Rhyssinae and Pimplinae (but with the exception of *Xanthopimpla*, see below). However, this grouping was not found in the nucleotide analyses (Figs 3c, d), and we recovered different (but partly weakly supported) relationships among the remaining subfamilies. The positions of Acaenitinae, Collyriinae, and Diplazontinae were highly inconsistent among our analyses (e.g., compare Figs 3a and 3e) and thus remain unclear, but the low support in this part of each tree precludes rejection of the hypothesis that they all branch off close to the base of a group containing the koinobiont, endoparasitoid pimpliform subfamilies . In contrast to previous suggestions , Diplazontinae and Orthocentrinae were never recovered as sister groups; instead, Diplazontinae were repeatedly resolved as sister to the remaining pimpliform subfamilies (Figs 1, 3c, 3e, 3f), even though this placement was again not always recovered. Instead of grouping with Diplazontinae, the Orthocentrinae formed a clade with Diacritinae and Cyloceriinae in almost all analyses, usually with good to even maximum support. A close relationship between Cyloceriinae and Orthocentrinae has long been suggested based on morphological and 28S evidence and some authors (e.g. Humala, 2007) maintain one subfamily for the cyloceriine and orthocentrine genera. Wahl (1990) defined a clade of Koinobiont endoparasitoids of Diptera comprising Cyloceriinae, Diplazontinae and Orthocentrinae based mainly on larval morphology: the distinctive fused hypostomal-stipital plate, not otherwise found in Ichneumonidae. Given these results and the lack of convincing apomorphies in adult morphology, it is possible the the hypostomal-stipital plate has evolved more than once in ichneumonid larvae feeding within Diptera larvae.

677 Diacritinae and Orthocentrinae have also been grouped together previously, especially
678by Perkins and Townes, who both included *Diacritus* Förster in their Plectiscinae (roughly
679equivalent to Orthocentrinae), even though Townes did so only "provisionally" and later
680classified Diacritinae as a separate tribe within an expanded Pimplinae that also included
681Rhyssinae and Poemeniinae. Morphologically, a close relationship between Orthocentrinae
682and Diacritinae is not without merit. *Diacritus* shares a fringe of dense setae along the inner
683side of the apex of the hind tibia with Orthocentrinae (even though it is somewhat less
684pronounced), and the first metasomal segment, with its elongate shape and fusion between
685sternite and tergite, is reminiscent of some orthocentrine genera such as *Proclitus* Förster,
686*Dialipsis* Förster and *Symplecis* Förster. The fact that both Orthocentrinae and Cyloceriinae
687are parasitoids of Diptera (Supplementary File S6) might provide a clue as to where to look
688for hosts of Diacritinae, for which no host data are available to date.

689 Wahl and Gauld suggested that the Acaenitinae consists of a grade of genera
690previously classified in the tribe Coleocentrini and the monophyletic former Acaenitini, but
691only referred to an unpublished, preliminary cladistic analysis to support this claim. It is a
692distinct possibility that the former grade is not only para- but indeed polyphyletic with respect
693to other pimpliform subfamilies. Morphological similarities between Acaenitinae and other
694subfamilies have been pointed out in the past, especially with Collyriinae, with which many
695acaenitine genera share the median tubercle on the clypeus, the propodeum lacking any
696transverse carinae, and the nervellus of the hind wing intercepted high up. In our analyses, the
697four genera of the former Acaenitini were always recovered together but almost never sharing
698a unique common ancestor with the sole representative of the former Coleocentrini included
699here, *Coleocentrus* (Figs 2 and 3). In the analyses of the transcriptome sequence data,
700*Coleocentrus* always clustered with *Collyria* (Fig. 1; no other acaenitine genera were
701included), and this sister-group relationship also appeared in some of the analyses of the full
702dataset (e.g., Fig. 3d). However, *Coleocentrus* clustered inconsistently with various
703pimpliform groups, depending on data type and analysis, and support for a separation of the
704two acaenitine groups was weak in most analyses. Furthermore, the Acaenitinae including
705*Coleocentrus* were recovered as monophyletic in some of the single-gene trees
706(Supplementary File S9). Thus, we await an increased taxon sampling of the *Coleocentrus*-
707group of genera (especially including specimens of *Procinetus* Förster) before suggesting any
708taxonomic changes.

709 The subfamily Pimplinae (excluding *Xanthopimpla*) was only recovered as
710 monophyletic in the analyses of the amino-acid alignment (Figs 2, 3a–d) and not those of the
711 nucleotide sequence data (Figs 3e–f). The placement of *Xanthopimpla* might have been
712 obscured by the long subtending branch of our exemplar species (*X. varimaculata* Cameron),
713 an issue that could be remedied in the future by a more extensive sampling of this species-rich
714 genus. *Xanthopimpla* shares at least two unique characters with *Lissopimpla* Kriechbaumer of
715 the Pimplini, namely the transversely divided clypeus and (with some exceptions in
716 *Xanthopimpla*) a transverse carina at the anterior end of the notauli (unfortunately, no
717 representatives of *Lissopimpla* were included in our study). Contrary to various authors ,
718 *Echthromorpha* Holmgren does not share the transversely divided clypeus and its mandibles,
719 although narrowed with the lower tooth very short, are not twisted as in *Lissopimpla* and
720 *Xanthopimpla*. *Xanthopimpla* also seems to be among the oldest extant pimpline genera when
721 considering the fossil record , which could support an early branching of this genus. Together
722 with *Lissopimpla*, *Xanthopimpla* might have to be included in a new tribe or even subfamily
723 in the future, but we await an analysis with more extensive taxon sampling before formalizing
724 such a change. As for the relationships among the other Pimplinae genera, it is interesting that
725 the results recovered here are strongly reminiscent of the tribal classification suggested by
726 Townes : he combined today's Delomeristini with the *Theronia*-group of genera and
727 *Pseudorhyssa* in the tribe Theroniini, but expressed uncertainty in both cases. Later cladistic
728 analyses based on morphological characters moved *Pseudorhyssa* to Poemeniinae and the
729 *Theronia*-group into Pimplini . Our analyses provide support for Townes's classification,
730 which was largely intuitive. The grouping of *Pseudorhyssa* in Delomeristini was recovered
731 with maximum support in all analyses of the full dataset (Fig. 2) and was consistent even
732 across single-gene trees (Supplementary File S11); we therefore move this genus to the tribe
733 Delomeristini (Pimplinae), with the result that the tribe Pseudorhyssini Wahl & Gauld, 1998,
734 becomes a junior synonym of Delomeristini Hellén, 1919; **syn nov.** We note that the
735 continuation of the epomia along the ventral edge of the pronotum in *Pseudorhyssa* is similar
736 to that found in *Delomerista* Förster and *Atractogaster* Kriechbaumer (the epomia is much
737 shorter in *Perithous*, the other genus of Delomeristini) and dissimilar to the more pronounced
738 bulging of the pronotum along the carina in Poemeniinae. Additionally, *Pseudorhyssa* has a
739 weak, ventrally placed glymma on the first metasomal tergite in common with other
740 Delomeristini and unlike Poemeniinae. The placement of the *Theronia* genus-group with
741 Delomeristini was not recovered in a majority of gene trees, and the current placement within
742 Pimplini also occurred in some gene trees. However, the group was consistently recovered as

743monophyletic here and is well-defined on morphological grounds ; thus, we resurrect the
744Theroniini as a tribe for the *Theronia* genus-group (stat. rev.), as was advocated by Carlson
745(1979).

746 Our results clearly placed *Hemiphanes* in the Ichneumoniformes-group of subfamilies,
747both in the concatenated analyses and in the vast majority of the gene trees but, on the basis of
748our taxon sampling, did not firmly suggest a subfamily placement. We here transfer
749*Hemiphanes* to Cryptinae. Whilst the morphology of *Hemiphanes* does not present any
750unambiguous synapomorphies with Cryptinae, it is a better fit than Orthocentrinae, as is borne
751out by our molecular results. Characters in support of this placement include the fused first
752metasomal sternite and tergite, absence of a glymma, anteriorly narrow and posteriorly much
753wider first tergite, and strongly sclerotized remaining tergites. There is a distinct sternaulus on
754the anterior 0.3 – 0.4 of the mesopleuron, hind wing vein M+Cu is straight and the nervellus
755(distal abscissa of Cu) is strong. Furthermore, fore wing vein 2m-cu has one bulla, the
756ovipositor has a nodus, and the male aedeagus is roughly ‘drop-shaped’ distally, not dorso-
757ventrally flattened as in Pimpliformes. In several respects, including the shape of the clypeus,
758large hypopygium, short first metasomal sternite, spiracle of the first tergite anterior to the
759mid-length, wing venation and general structure, *Hemiphanes* is surprisingly similar to
760*Sphecophaga* Westwood, a cryptine parasitoid of eusocial vespine wasps; the two key out to
761the same couplet in a recent key to ichneumonid subfamilies . Features such as the twisted
762mandibles and fringe of setae on the inner aspect of the hind tibia do not strongly support the
763placement of *Hemiphanes* in the Orthocentrinae, as the mandibles are more robust and fringe
764less dense than in most orthocentrines. The sum of characters suggest that *Hemiphanes*,
765particularly the wing venation and reduced propodeal carinae, suggest that *Hemiphanes*
766belongs in Cryptinae rather than Phygadeuontinae, and the pronounced, triangular expansions
767of the metanotum are a feature of the tribe Aptesini rather than Cryptini, so we place
768*Hemiphanes* in Aptesini, despite its overall similarity with *Sphecophaga*.

769Rapid radiation of Pimpliformes

770 Rapid radiations have been suggested in several groups of parasitoid wasps and are
771mostly attributed to the formation of new ecological niches through host switches. A rapid
772radiation after host switching is possibly the case for the microgastroid subfamilies of the
773Braconidae, which changed to lepidopteran hosts and at the same time acquired poly-DNA
774viruses to overcome their hosts' immune systems . However, previous studies inferring rapid
775radiations were based on the analysis of small sets of genes, usually no more than three or

four, which could explain the low resolution in the phylogenetic inferences. In Chalcidoidea, a recent transcriptome study based on more than 3,200 genes suggested a rapid radiation of some of its families during the late Cretaceous, although the inferred branch lengths of their dated tree still mostly span about 5–10 million years and are, thus, not that rapid after all. Our study might therefore be viewed as the most convincing example to date of a rapid radiation in parasitoid wasps, supported by both considerable gene sampling and very short inferred branch lengths (Fig. 2). As a very rough estimate, assuming the most recent common ancestor of Pimpliformes existed about 66 million years ago (see below) and using average branch lengths as clock lengths, the shortest branches at the base of the radiation would be less than one million years long; but this result should be confirmed by a properly calibrated molecular clock analysis.

Our ancestral-state reconstructions inferred at least one switch between idiobiosis and koinobiosis during the most rapid phase of the radiation (Fig. 3). In addition, Diptera were acquired as hosts by the subfamilies Diplazontinae, Orthocentrinae, and Cyloceriinae (and possibly Diacritinae), according to our trees probably in two separate events (but see previous section where we discuss that we cannot exclude that they are indeed monophyletic, which would imply a single switch). Diptera are rather rare as hosts for Ichneumonidae, and the effect of this host switch on diversification rates and patterns deserves further study. The same is true for the switches between parasitoid strategies, including between ecto- and endoparasitoidism. We here refrain from a detailed discussion of the implications of our results, given the inconsistency between analyses (Fig. 3). A better-resolved tree is needed to address such questions.

The timing of the rapid radiation of the pimpliform subfamilies remains unclear as no dated tree for the group is yet available. However, the fossil record indicates that it might have taken place around or just after the Cretaceous-Paleogene boundary. Fossil ichneumonids from the Cretaceous are all classified in now extinct subfamilies, with the exception of a species from Canadian amber that was tentatively placed in Labeninae, a group consistently branching earlier than Pimpliformes in our analyses. On the other hand, several clear representatives of the subfamilies Acaenitinae, Orthocentrinae, Pimplinae and Rhyssinae are known from the upper and middle Eocene. If a dated phylogenetic analysis confirms the timing of the pimpliform radiation as directly after the K-Pg boundary, this group's history can be compared to other rapid radiations that have presumably taken place around the same time, such as that of Neoaves, placental mammals and legumes.

809 *Do we need more genes, more taxa, or better methods?*

810 Rapid radiations that happened in the deep past constitute some of the most difficult
811 phylogenetic problems. The low probability of any substitution happening along the short
812 basal branches, combined with the risk that such true signal is masked by subsequent changes
813 along the typically long branches leading from there, might mean that very large amounts of
814 data need to be generated in order to observe sufficient signal for resolving deep, rapid
815 radiations. Biases in the data arising from model misspecification, non-random patterns of
816 missing data, or among-lineage rate variation and associated effects such as long-branch
817 attraction, might easily obscure any true phylogenetic signal from such radiations.
818 Consequently, several studies targeting old radiations using genome-scale datasets have failed
819 to completely resolve the underlying species tree, and others retrieved conflicting hypotheses
820 depending on dataset or analytical methods.

821 Is it possible to distinguish between a lack of signal due to an extremely rapid
822 radiation (or in other words, a truly hard polytomy), insufficient gene or taxon sampling, and
823 systematic biases in a dataset? Analyses of subsets of a dataset can, to a certain extent,
824 exclude under-sampling as a major disruptive factor. Our phylogenetic inferences based on
825 gene sets of different sizes indicate a slow-down in the increase of mean clade support with
826 increasing gene sampling and indications of a levelling-off at an average bootstrap per node
827 of about 0.9, even though it is not entirely clear if a further increase might still be possible if
828 more than 93 genes were sampled (Fig. 4).

829 It also remains unclear whether our approach to marker choice has in fact improved
830 the inference of the phylogeny, given that neither of the criteria clocklikeness, number of sites
831 at near-optimal rates, or single-gene bootstrap support clearly outperformed mere gene length
832 in the analyses of subsets of genes (Fig. 4). However, it is possible that the initial choice of
833 markers from the transcriptome gene-set already eliminated most of the inferior genes and,
834 thus, left us with insufficient power to detect any remaining effect of these predictors. The
835 efficiency of some of the criteria we applied are also not beyond doubt. For instance, clade
836 support values for certain clades might be high despite very low resolution in other parts of
837 the tree, especially if there is a difference in node depth. Using sums of clade support might
838 therefore favour genes that resolve the more numerous, shallow nodes well, while not
839 accounting sufficiently for the deeper nodes. Furthermore, we used a combined approach with
840 the sequential application of several different criteria for choosing genes, and the relative
841 performance and combinability of these criteria still needs to be evaluated.

842 Tests for the amount and extent of taxon sampling that is necessary to resolve a deep
843 radiation are even more difficult to perform, as the addition of taxa always changes the nature
844 of the phylogenetic problem. In our case, the trees derived from the 723-gene transcriptome
845 dataset of 26 taxa was fully resolved in terms of maximum bootstrap support in all analyses,
846 while the combined hybrid capture data of 93 genes of 84 taxa exposed the weakly-resolved
847 rapid radiation at the base of Pimpliformes. The larger dataset included the additional
848 subfamilies Cylloceriinae, Diacritinae, and Orthocentrinae, and several additional tribes.
849 Thus, it included additional nodes which were difficult to resolve. The most striking
850 topological difference between the trees was the placement of Diplazontinae: breaking down
851 the branch leading to *Syrphophilus* Dasch, the only diplazontine included in the transcriptome
852 dataset, obviously had the effect of moving this subfamily further up the pimpliform tree
853 (compare Figs 2 and 3). Consideration of phylogenetic informativeness led to the conclusion
854 that targeted taxon sampling should focus on species that would branch off as closely to the
855 unresolved nodes as possible. In our case, a promising approach would be to sample multiple
856 genera per subfamily, such as in Diacritinae and Collyriinae, but also focusing on more
857 basally-branching members of other subfamilies such as Acaenitinae. While such an approach
858 of course cannot increase the length of the most basal branches of the pimpliform radiation, it
859 breaks down their long subtending branches and can therefore ameliorate long-branch
860 attraction phenomena that could obscure the true phylogenetic signal.

861 Systematic biases such as nucleotide composition biases and other mismatches of the
862 substitution model can be exposed by analytical methods, for example, the test for uneven
863 nucleotide composition performed here, but also by more advanced methods falling into the
864 class of model adequacy testing. In most cases, such biases may be datatype-specific and will
865 thus translate into conflicts between different dataset types, such as amino acids versus
866 nucleotides or all three versus the first two codon positions. Such conflicts were apparent
867 especially in our transcriptome sequence analyses, while the complete data typically exhibited
868 low support at those nodes that were in conflict between analyses. The absence of supported
869 conflict between data types might indicate that substitution-model misspecification was not
870 the major factor preventing resolution of the pimpliform radiation. Also, in contrast to
871 previous analyses that recovered strongly supported conflict among the single-gene trees,
872 those here were mostly unresolved at the base of Pimpliformes, which indicates that gene-
873 tree/species-tree analyses would not yield improved resolution. Nevertheless, better
874 evolutionary models might facilitate the extraction of what little signal is left in sequence data
875 from the early and rapid pimpliform radiation. Another alternative is the analysis of different

876data types that show comparatively low levels of homoplasy, for instance certain ecological or
877morphological characters or characters pertaining to genome morphology .

878 In conclusion, we postulate here one of the most rapid radiations uncovered to date in
879parasitoid wasps and speculate that the most promising approach to improve its resolution
880probably lies in increased taxon sampling, even though the sampling of additional markers or
881even other data types remains an alternative. The molecular resources and analysis pipeline
882we present here represent a significant step forward in phylogenetics of ichneumonids and
883beyond and will in the future hopefully allow a more robust inference of the relationships
884among these wasps and of the evolutionary history of their varied parasitoid ecologies.

885

886Acknowledgments

887 We are very grateful to everyone who provided rare specimens for this study: Eric
888Chapman (University of Kentucky, U.S.A.), the Swedish Malaise Trap Project
889(www.stationlinne.se, Sweden), Masato Ito (Kobe University, Japan), Ilari Sääksjärvi
890(University of Turku, Finland), Stanislav Korenko (Czech University of Life Sciences
891Prague), Philippe Louâpre (Université de Bourgogne, France), and Marc Neumann
892(Walterswil SO, Switzerland). Muriel Fragnière (University of Bern, Switzerland) helped
893with the planning of the transcriptome part of this study, prepared the RNA libraries and
894conducted the transcriptome sequencing. We also thank Kathy Saint and Tessa Bradford
895(University of Adelaide, Australia) for helpful advice on the hybrid-enrichment lab work.
896Hugh Robertson (University of Illinois, U.S.A.) provided us with access to the genome of
897*Microplitis demolitor* before its release for assessing intron-exon boundaries for bait design.

898 This study was funded by the Swiss National Science Foundation (grant
899PZ00P3_154791 to SK).

900

901References

902
903Andrews S. (2014) FastQC version 0.11.2. Cambridge, Babraham Institute.
904Askew R.R., Shaw M.R. (1986) Parasitoid communities: Their size, structure and
905development. In: Waage JK, Greathead D editors. *Insect Parasitoids*. London, Academic
906Press, p. 225–264.
907Austin A.D. (1985) The function of spider egg sacs in relation to parasitoids and predators,
908with special reference to the Australian fauna. *Journal of Natural History*, **19**, 359–376.

909Belshaw R., Fitton M.G., Herniou E., Gimeno C., Quicke D.L.J. (1998) A phylogenetic
910reconstruction of the Ichneumonoidea (Hymenoptera) based on the D2 variable region of
91128S rivosomal RNA. *Systematic Entomology*, **23**, 109–123.

912Bergsten J., Bilton D.T., Fujisawa T., Elliott M., Monaghan M.T., Balke M., Hendrich L.,
913Geijer J., Herrmann J., Foster G.N., Ribera I., Nilsson A.N., Barraclough T.G., Vogler A.P.
914(2012) The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*,
915**61**, 851–869.

916Bolger A.M., Lohse M., Usadel B. (2014) Trimmomatic: A flexible trimmer for Illumina
917Sequence Data. *Bioinformatics*, **30**, 2114–2120.

918Bragg J.G., Potter S., Bi K., Moritz C. (2016) Exon capture phylogenomics: efficacy across
919scales of divergence. *Molecular Ecology Resources*, **16**, 1059–1068.

920Broad G.R., Shaw M.R., Fitton M.G. (2018) *The ichneumonid wasps of Britain and Ireland*
921(Hymenoptera: Ichneumonidae). Royal Entomological Society, Telford.

922Brown J.M. (2014) Detection of implausible phylogenetic inferences using posterior predictive
923assessment of model fit. *Systematic Biology*, **63**, 334–348.

924Carlson, R.W. (1979) Family Ichneumonidae. In: Krombein, K.V., Hurt, P.D., Smith, D.R., Burks, B.D., editors.
925Catalog of Hymenoptera in America north of Mexico. Washington. Pp. 315-740.

926Collins T.M., Fedrigo O., Naylor G.J.P. (2005) Choosing the best genes for the job: the case
927for stationary genes in genome-scale phylogenetics. *Systematic Biology*, **54**, 493–500.

928Dibb N.J., Newman A.J. (1989) Evidence that introns arose at proto-splice sites. *The EMBO*
929*Journal*, **8**, 2015–2021.

930dos Reis M., Inoue J., Hasegawa M., Asher R.J., Donoghue P.C.J., Yang Z. (2012)
931Phylogenomic datasets provide both precision and accuracy in estimating the timescale of
932placental mammal phylogeny. *Proceedings of the Royal Society of London Series B-*
933*Biological Sciences*, **279**, 3491–3500.

934Dowton M., Cameron S., Dowavac J.I., Austin A.D., Whiting M.F. (2009) Characterization of
93567 mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that
936mitochondrial tRNA gene position is selectively neutral. *Molecular Biology and Evolution*, **26**,
9371607–1617.

938Doyle V.P., Young R.E., Naylor G.J.P., Brown J.M. (2015) Can we identify genes with
939increased phylogenetic reliability? *Systematic Biology*, **64**, 824–837.

940Duchêne D., Duchêne S., Ho S.Y.W. (2018) PhyloMAAd: Efficient assessment of
941phylogenomic model adequacy. *Bioinformatics*, **bty103**.

942Edwards S.V., Liu L., Pearl D.K. (2007) High-resolution species trees without concatenation.
943*Proceedings of the National Academy of Sciences of the United States of America*, **104**,
9445936–5941.

945Eggleton P. (1989) The phylogeny and evolutionary biology of the Pimplinae (Hymenoptera:
946Ichneumonidae). Department of Pure and Applied Biology. London, Imperial College, p. 295.

947Ericson P.G.P., Klopstein S., Irestedt M., Nguyen J.M.T., Nylander J.A.A. (2014) Dating the
948diversification of the major lineages of Passeriformes (Aves). *BMC Evolutionary Biology*, **14**,
9498.

950Fischer M., Steel M.A. (2009) Sequence length bounds for resolving a deep phylogenetic
951divergence. *Journal of Theoretical Biology*, **256**, 247–252.

952Gatesy J., DeSalle R., Wahlberg N. (2007) How many genes should a systematist sample?
953Conflicting insights from a phylogenomic matrix characterized by replicated incongruence.
954*Systematic Biology*, **56**, 355–363.

955Gauld I.D. (1988) Evolutionary patterns of host utilization by ichneumonid parasitoids
956(Hymenoptera: Ichneumonidae and Braconidae). *Biological Journal of the Linnean Society*,
957**35**, 351–377.

958Gauld I.D. (1991) The Ichneumonidae of Costa Rica, 1. *Memoirs of the American*
959*Entomological Institute*, **47**, 1–589.

960Gauld I.D., Dubois J. (2006) Phylogeny of the *Polysphincta* group of genera (Hymenoptera:
961Ichneumonidae; Pimplinae): a taxonomic revision of spider ectoparasitoids. *Systematic*
962*Entomology*, **31**, 529-564.

963Gauld I.D., Mound L.A. (1982) Homoplasy and the delineation of holophyletic genera in some
964insect groups. *Systematic Entomology*, **7**, 73–86.

965Gauld I.D., Sithole R., Gómez J.U., Godoy C. (2002a) The Ichneumonidae of Costa Rica, 4.
966*Memoirs of the American Entomological Institute*, **66**, 1–768.

967Gauld I.D., Wahl D.B., Broad G.R. (2002b) The suprageneric groups of the Pimplinae
968(Hymenoptera: Ichneumonidae): a cladistic re-evaluation and evolutionary biological study.
969*Zoological Journal of the Linnean Society*, **136**, 421–485.

970Gauld I.D., Ward S., Mallet V. (2000) The Ichneumonidae of Costa Rica, 3. Introduction and
971keys to species of the subfamilies Brachycyrtinae, Cremastinae, Labeninae and Oxytorinae,
972with an appendix on the Anomaloniinae. *Memoirs of the American Entomological Institute*, **63**,
9731–453.

974Geuten K., Massingham T., Darius P., Smets E., Goldman N. (2007) Experimental design
975criteria in phylogenetics: where to add taxa. *Systematic Biology*, **56**, 609–622.

976Glenn T.C., Nilsen R., Kieran T.J., Finger J.W., Pierson T.W., Bentley K.E., Hoffberg S.,
977Louha S., Garcia-De-Leon F.J., del Rio Portilla M.A., Reed K., Anderson J.L., Meece J.K.,
978Aggery S., Rekaya R., Alabady M., Belanger M., Winker K., Faircloth B.C. (2016)
979Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries
980(iTru & iNext).

981Goldman N. (1998) Phylogenetic information and experimental design in molecular
982systematics. *Proceedings of the Royal Society of London Series B-Biological Sciences*, **265**,
9831779–1786.

984Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan
985L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di
986Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. (2011) Full-
987length transcriptome assembly from RNA-Seq data without a reference genome. *Nature*
988*Biotechnology*, **29**, 644–654.

989Hartig G., Peters R.S., Borner J., Etzbauer C., Misof B., Niehuis O. (2012) Oligonucleotide
990primers for targeted amplification of single-copy nuclear genes in Apocritan Hymenoptera.
991*PLoS One*, **7**, e39826.

992Hasegawa M., Kishino H., Yano T. (1985) Dating of the human-ape splitting by a molecular
993clock of mitochondrial DNA. *Journal of Molecular Evolution*, **22**, 160–174.

994Heath T.A., Hedtke S.M., Hillis D.M. (2008) Taxon sampling and the accuracy of
995phylogenetic analyses. *Journal of Systematics and Evolution*, **46**, 239–257.

996Hedtke S.M., Townsend T.M., Hillis D.M. (2006) Resolution of phylogenetic conflict in large
997data sets by increased taxon sampling. *Systematic Biology*, **55**, 522–529.

998Henricson A., Forslund K., Sonnhammer E.L.L. (2010) Orthology confers intron position
999conservation. *BMC Genomics*, **11**, 412.

1000Huang X., Madan A. (1999) CAP3: A DNA Sequence Assembly Program. *Genome*
1001*Research*, **9**, 868–877.

1002Hugall A., O'Hara T., Hunjan S., Nilsen R., Moussali A. (2015) An exon-capture system for
1003the entire class Ophiuroidea. *Molecular Biology and Evolution*, **33**, 281–294.

1004Humala, A.E. (2007) To the system of the subfamily Orthocentrinae s. l. (Hymenoptera, Ichneumonidae). In:
1005Rasnitsyn, A.P., Gokhman, V.E., editors. [Studies on hymenopterous insects. Collection of scientific papers.]
1006(In Russian with English summaries). Moscow: KMK Scientific Press Ltd. Pp. 76-85.

1007i5K Consortium. (2013) The i5K initiative: advancing arthropod genomics for knowledge,
1008human health, agriculture, and the environment. *Journal of Heredity*, **104**, 595–600.

1009IUCN. (2017) IUCN Red List of Threatened Species, version 2017.3. Gland, Switzerland,
1010IUCN.

1011Jarvis E.D., Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y.W., Faircloth B.C.,
1012Nabholz B., Howard J.T., Suh A., Weber C.C., da Fonseca R.R., Li J.W., Zhang F., Li H.,
1013Zhou L., Narula N., Liu L., Ganapathy G., Boussau B., Bayzid M.S., Zavidovych V.,
1014Subramanian S., Gabaldon T., Capella-Gutierrez S., Huerta-Cepas J., Rekepalli B., Munch
1015K., Schierup M., Lindow B., Warren W.C., Ray D., Green R.E., Bruford M.W., Zhan X.J.,
1016Dixon A., Li S.B., Li N., Huang Y.H., Derryberry E.P., Bertelsen M.F., Sheldon F.H.,
1017Brumfield R.T., Mello C.V., Lovell P.V., Wirthlin M., Schneider M.P.C., Prosdocimi F.,
1018Samaniego J.A., Velazquez A.M.V., Alfaro-Nunez A., Campos P.F., Petersen B., Sicheritz-
1019Ponten T., Pas A., Bailey T., Scofield P., Bunce M., Lambert D.M., Zhou Q., Perelman P.,
1020Driskell A.C., Shapiro B., Xiong Z.J., Zeng Y.L., Liu S.P., Li Z.Y., Liu B.H., Wu K., Xiao J.,
1021Yinqi X., Zheng Q.M., Zhang Y., Yang H.M., Wang J., Smeds L., Rheindt F.E., Braun M.,

1022Fjeldsa J., Orlando L., Barker F.K., Jonsson K.A., Johnson W., Koepfli K.P., O'Brien S.,
1023Haussler D., Ryder O.A., Rahbek C., Willerslev E., Graves G.R., Glenn T.C., McCormack J.,
1024Burt D., Ellegren H., Alstrom P., Edwards S.V., Stamatakis A., Mindell D.P., Cracraft J.,
1025Braun E.L., Warnow T., Jun W., Gilbert M.T.P., Zhang G.J. (2014) Whole-genome analyses
1026resolve early branches in the tree of life of modern birds. *Science*, **346**, 1320–1331.
1027Katoh K., Standley D.M. (2013) MAFFT multiple sequence alignment software version 7:
1028improvements in performance and usability. *Molecular Biology and Evolution*, **30**, 772–780.
1029Klopfstein S. (2014) Revision of the Western Palaearctic Diplazontinae (Hymenoptera,
1030Ichneumonidae). *Zootaxa*, **3801**, 1–143.
1031Klopfstein S., Kropf C., Baur H. (2016) *Wolbachia* endosymbionts distort DNA barcoding in
1032the parasitoid wasp genus *Diplazon* (Hymenoptera: Ichneumonidae). *Zoological Journal of*
1033*the Linnean Society*, **177**, 541–557.
1034Klopfstein S., Kropf C., Quicke D.L.J. (2010) An evaluation of phylogenetic informativeness
1035profiles and the molecular phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae).
1036*Systematic Biology*, **59**, 226–241.
1037Klopfstein S., Massingham T., Goldman N. (2017) More on the best evolutionary rate for
1038phylogenetic analysis. *Systematic Biology*, **66**, 769–785.
1039Klopfstein S., Quicke D.L.J., Kropf C., Frick H. (2011) Molecular and morphological
1040phylogeny of Diplazontinae (Hymenoptera, Ichneumonidae). *Zoologica Scripta*, **40**, 379–402.
1041Klopfstein S., Ronquist F. (2013) Convergent intron gains in hymenopteran elongation factor-
10421 α . *Molecular Phylogenetics and Evolution*, **67**, 266–276.
1043Koenen E.J.M., de Vos J.M., Atchison G.W., Simon M.F., Schrire B.D., de Souza E.R., de
1044Queiroz L.P., Hughes C.E. (2013) Exploring the tempo of species diversification in legumes.
1045*South African Journal of Botany*, **89**, 19–30.
1046Kopylov D.S. (2009) A new subfamily of ichneumons from the lower Cretaceous of
1047Transbaikalia and Mongolia (Insecta: Hymenoptera: Ichneumonidae). (in Russian).
1048*Paleontologicheskii Zhurnal*, **1**, 76–85.
1049Kopylov D.S. (2010) A new subfamily of ichneumon wasps (Insecta: Hymenoptera:
1050Ichneumonidae) from the Upper Cretaceous of the Russian Far East. *Paleontological*
1051*Journal*, **4**, 59–69.
1052Kopylov D.S. (2012) New species of Praeichneumonidae (Hymenoptera, Ichneumonoidea)
1053from the Lower Cretaceous of Transbaikalia. *Paleontological Journal*, **46**, 66–72.
1054Ksepka D.T., Phillips M.J. (2015) Avian diversification patterns across the K-Pg boundary:
1055influence of calibrations, datasets, and model misspecification. *Annals of the Missouri*
1056*Botanical Garden*, **100**, 300–328.
1057Lanfear R., Frandsen P.B., Wright A.M., Senfeld T., Calcott B. (2016) PartitionFinder 2: new
1058methods for selecting partitioned models of evolution for molecular and morphological
1059phylogenetic analyses. *Molecular Biology and Evolution*, **34**, 772–773.
1060Lartillot N., Philippe H. (2004) A Bayesian mixture model for across-site heterogeneities in
1061the amino-acid replacement process. *Molecular Biology and Evolution*, **21**, 1095–1109.
1062Lepage T., Bryant D., Philippe H., Lartillot N. (2007) A general comparison of relaxed
1063molecular clock models. *Molecular Biology and Evolution*, **24**, 2669–2680.
1064Liu L., Pearl D.K., Brumfield R.T., Edwards S.V. (2008) Estimating species trees using
1065multiple-allele DNA sequence data. *Evolution*, **62**, 2080–2091.
1066Louca S., Doebeli M. (2018) Efficient comparative phylogenetics on large trees.
1067*Bioinformatics*, **34**, 1053–1055.
1068Mamanova L., Coffey A.J., Scott C.E., Kozarewa I., Turner E.H., Kumar A., Howard E.,
1069Shendure J., Turner D.J. (2010) Target-enrichment strategies for next-generation
1070sequencing. *Nature Methods*, **7**, 111–118.
1071Martin M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing
1072reads. *EMBnet Journal*, **17**, 10–12.
1073Mayer C., Sann M., Donath A., Meixner M., Podsiadlowski L., Peters R.S., Petersen M.,
1074Meusemann K., Lierse K., Waagele J.W., Misof B., Bleidorn C., Ohl M., Niehuis O. (2016)
1075BaitFisher: a software package of multispecies target DNA enrichment probe design.
1076*Molecular Biology and Evolution*, **33**, 1875–1886.

1077McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T.
1078(2013a) A phylogeny of birds based on over 1,500 loci collected by target enrichment and
1079high-throughput sequencing. *PLoS One*, **8**, e54858.

1080McCormack J.E., Hird S.M., Zellmer A.J., Carstens B.C., Brumfield R.T. (2013b) Applications
1081of next-generation sequencing to phylogeography and phylogenetics. *Molecular*
1082*Phylogenetics and Evolution*, **66**, 526–538.

1083McKellar R.C., Kopylov D.S., Engel M.S. (2013) Ichneumonidae (Insecta: Hymenoptera) in
1084Canadian Late Cretaceous amber. *Fossil Record*, **16**, 217–227.

1085Meusemann K., von Reumont B.M., Simon S., Roeding F., Strauss S., Kück P., Ebersberger
1086I., Walz M., Pass G., Breuers S., Achter V., Von Haeseler A., Burmester T., Hadrys H.,
1087Waegele J.W., Misof B. (2010) A phylogenomic approach to resolve the arthropod tree of life.
1088*Molecular Biology and Evolution*, **27**, 2451–2464.

1089Meyer M., Kircher M. (2010) Illumina sequencing library preparation for highly multiplexed
1090target capture and sequencing. *Cold Spring Harbor Protocols*, **2010**, 1–20.

1091Misof B., Liu S.L., Meusemann K., Peters R.S., Donath A., Mayer C., Frandsen P.B., Ware
1092J., Flouri T., Beutel R.G., Niehuis O., Petersen M., Izquierdo-Carrasco F., Wappler T., Rust
1093J., Aberer A.J., Aspöck U., Aspöck H., Bartel D., Blanke A., Berger S., Böhm A., Buckley
1094T.R., Calcott B., Chen J.Q., Friedrich F., Fukui M., Fujita M., Greve C., Grobe P., Gu S.C.,
1095Huang Y., Jermiin L.S., Kawahara A.Y., Krogmann L., Kubiak M., Lanfear R., Letsch H., Li
1096Y.Y., Li Z.Y., Li J.G., Lu H.R., Machida R., Mashimo Y., Kapli P., McKenna D.D., Meng G.L.,
1097Nakagaki Y., Navarrete-Heredia J.L., Ott M., Ou Y.X., Pass G., Podsiadlowski L., Pohl H.,
1098von Reumont B.M., Schütte K., Sekiya K., Shimizu S., Slipinski A., Stamatakis A., Song
1099W.H., Su X., Szucsich N.U., Tan M.H., Tan X.M., Tang M., Tang J.B., Timelthaler G.,
1100Tomizuka S., Trautwein M., Tong X.L., Uchifune T., Walz M.G., Wiegmann B.M., Wilbrandt
1101J., Wipfler B., Wong T.K.F., Wu Q., Wu G.X., Xie Y.L., Yang S.Z., Yang Q., Yeates D.K.,
1102Yoshizawa K., Zhang Q., Zhang R., Zhang W.W., Zhang Y.H., Zhao J., Zhou C.R., Zhou
1103L.L., Ziesmann T., Zou S.J., Li Y.R., Xu X., Zhang Y., Yang H.M., Wang J., Wang J., Kjer
1104K.M., Zhou X. (2014) Phylogenomics resolves the timing and pattern of insect evolution.
1105*Science*, **346**, 763–767.

1106Murphy N.P., Banks J.C., Whitfield J.B., Austin A.D. (2008) Phylogeny of the parasitic
1107microgastroid subfamilies (Hymenoptera: Braconidae) based on sequence data from seven
1108genes, with an improved time estimate of the origin of the lineage. *Molecular Phylogenetics*
1109*and Evolution*, **47**, 378–395.

1110Murphy N.P., Carey D., Castro L.R., Dowton M., Austin A.D. (2007) Phylogeny of the
1111platygastroid wasps (Hymenoptera) based on sequences from the 18S rRNA, 28S rRNA and
1112cytochrome oxidase I genes: implications for the evolution of the ovipositor system and host
1113relationships. *Biological Journal of the Linnean Society*, **91**, 653–669.

1114Narendran T.C. (2001) Parasitic Hymenoptera and biological control. In: Upadhyay RK,
1115Mukerji KG, Chamola BP editors. Biocontrol potential and its exploitation in sustainable
1116agriculture. New York, Kluwer Academic / Plenum Publishers.

1117Nicholls J.A., Challis R.J., Mutun S., Stone G.N. (2012) Mitochondrial barcodes are
1118diagnostic of shared refugia but not species in hybridizing oak gallwasps. *Molecular Ecology*,
1119**21**, 4051–4062.

1120Nylander J.A.A., Ronquist F., Huelsenbeck J.P., Nieves-Aldrey J.L. (2004) Bayesian
1121phylogenetic analysis of combined data. *Systematic Biology*, **53**, 47–67.

1122Olsen G.L. (1996) DNArates - maximum likelihood site rate calculation. Published by the
1123author.

1124Paradis E., Claude J., Strimmer K. (2004) APE: analyses of phylogenetics and evolution in R
1125language. *Bioinformatics*, **20**, 289–290.

1126Perkins J.F. (1940) Notes on the synonymy of some genera of European Pimplinae (s.l.)
1127(Hym. Ichneumonidae). *The Entomologist*, **73**, 54–56.

1128Peters R.S., Krogmann L., Mayer C., Donath A., Gunkel S., Meusemann K., Kozlov A.,
1129Podsiadlowski L., Petersen M., Lanfear R., Diez P.A., Heraty J.M., Kjer K.M., Klopstein S.,
1130Meier R., Polidori C., Schmitt T., Liu S.L., Zhou X., Wappler T., Rust J., Misof B., Niehuis O.
1131(2017) Evolutionary history of the Hymenoptera. *Current Biology*, **27**, 1013–1018.

1132 Peters R.S., Niehuis O., Gunkel S., Bläser M., Mayer C., Podsiadlowski L., Kozlov A.,
 1133 Donath A., Van Noort S., Liu S., Zhou X., Misof B., Heraty J.M., Krogmann L. (2018)
 1134 Transcriptome sequence-based phylogeny of chalcidoid wasps (Hymenoptera: Chalcidoidea)
 1135 reveals a history of rapid radiations, convergence, and evolutionary success. *Molecular*
 1136 *Phylogenetics and Evolution*, **120**, 286–296.
 1137 Petersen M., Meusemann K., Donath A., Dowling D., Liu S., Peters R.S., Podsiadlowski L.,
 1138 Vasilikopoulos A., Zhou X., Misof B., Niehuis O. (2017) Orthograph: a versatile tool for
 1139 mapping coding nucleotide sequences to clusters of orthologous genes. *BMC Bioinformatics*,
 1140 **18**, 111.
 1141 Pham N.T., Broad G.R., Waagele J.W. (2013) A review of the *Theronia* genus-group
 1142 (Hymenoptera: Ichneumonidae: Pimplinae) from Vietnam with descriptions of five new
 1143 species. *Journal of Natural History*, **47**, 1501–1538.
 1144 Philippe H., Brinkmann H., Lavrov D.V., Littlewood D.T.J., Manuel M., Wörheide G., Baurain
 1145 D. (2011) Resolving difficult phylogenetic questions: why more sequences are not enough.
 1146 *PLoS Biology*, **9**, e1000602.
 1147 Phillips M.J., Delsuc F., Penny D. (2004) Genome-scale phylogeny and the detection of
 1148 systematic biases. *Molecular Biology and Evolution*, **21**, 1455–1458.
 1149 Prum R.O., Berv J.S., Dornburg A., Field D.J., Townsend J.P., Moriarty Lemmon E.,
 1150 Lemmon A.R. (2015) A comprehensive phylogeny of birds (Aves) using targeted next-
 1151 generation DNA sequencing. *Nature*, **526**, 569–573.
 1152 Quicke D.L.J. (2015) Phylogeny and systematics of the Ichneumonidae. In: Quicke DLJ
 1153 editor. The braconid and ichneumonid parasitoid wasps: biology, systematics, evolution and
 1154 ecology. Chichester, UK, John Wiley & Sons, Ltd.
 1155 Quicke D.L.J., Fitton M.G., Notton D.G., Belshaw R., Broad G.R., Dolphin K. (2000)
 1156 Phylogeny of the Ichneumonidae (Hymenoptera). A simultaneous molecular and
 1157 morphological analysis. In: Austin AD, Downton M editors. Hymenoptera: Evolution,
 1158 Biodiversity and Biological Control. Canberra, Australia, CSIRO, p. 74–83.
 1159 Quicke D.L.J., Laurence N.M., Fitton M.G., Broad G.R. (2009) A thousand and one wasps: a
 1160 28S rDNA and morphological phylogeny of the Ichneumonidae (Insecta: Hymenoptera) with
 1161 an investigation into alignment parameter space and elision. *Journal of Natural History*, **43**,
 1162 1305–1421.
 1163 Quicke D.L.J., Lopez-Vaamonde C., Belshaw R. (1999) The basal Ichneumonidae (Insecta,
 1164 Hymenoptera): 28S D2 rDNA considerations of the Brachycyrtinae, Labeninae,
 1165 Paxylommatainae and Xoridinae. *Zoologica Scripta*, **28**, 203–210.
 1166 R Core Team. (2014) R: A language and environment for statistical computing. Vienna,
 1167 Austria, R Foundation for Statistical Computing.
 1168 Rago A., Gilbert D.G., Choi J.H., Sackton T.B., Wang X., Kelkar Y.D., Werren J.H.,
 1169 Colbourne J.K. (2016) OGS2: genome re-annotation of the jewel wasp *Nasonia vitripennis*.
 1170 *BMC Genomics*, **17**, 678.
 1171 Reddy S., Kimball R.T., Pandey A., Hosner P.A., Braun M.J., Hackett S.J., Han K.-L.,
 1172 Harshman J., Huddleston C.J., Kingston S., Marks B.D., Miglia K.J.M., W.S., Sheldon F.H.,
 1173 Witt C.C., Yuri T., Braun E.L. (2017) Why do phylogenomic data sets yield conflicting trees?
 1174 Data type influences the avian tree of life more than taxon sampling. *Systematic Biology*, **66**,
 1175 857–879.
 1176 Regier J.C., Shultz J.W., Ganley A.R.D., Hussey A., Shi D., Ball B., Zwick A., Stajich J.E.,
 1177 Cummings M.P., Martin J.W., Cunningham C.W. (2008) Resolving arthropod phylogeny:
 1178 exploring phylogenetic signal within 41 kb of protein-coding nuclear gene sequence.
 1179 *Systematic Biology*, **57**, 920–938.
 1180 Rodríguez-Ezpeleta N., Brinkmann H., Roure B., Lartillot N., Lang B.F., Philippe H. (2007)
 1181 Detecting and overcoming systematic errors in genome-scale phylogenies. *Systematic*
 1182 *Biology*, **56**, 389–399.
 1183 Rokas A., Williams B.L., King N., Carroll S.B. (2003) Genome-scale approaches to resolving
 1184 incongruence in molecular phylogenies. *Nature*, **425**, 798–804.
 1185 Ronquist F., Klopstein S., Vilhelmsen L., Schulmeister S., Murray D.L., Rasnitsyn A.P.
 1186 (2012a) A total-evidence approach to dating with fossils, applied to the early radiation of the
 1187 Hymenoptera. *Systematic Biology*, **61**, 973–999.

1188Ronquist F., Teslenko M., Van der Mark P., Ayres D.L., Darling A., Höhna S., Larget B., Liu
 1189L., Suchard M.A., Huelsenbeck J.P. (2012b) MrBayes 3.2: Efficient Bayesian phylogenetic
 1190inference and model choice across a large model space. *Systematic Biology*, **61**, 539–542.
 1191Rousse P., Quicke D.L.J., Matthee C.A., Lefevre P., Van Noort S. (2016) A molecular and
 1192morphological reassessment of the phylogeny of the subfamily Ophioninae (Hymenoptera:
 1193Ichneumonidae). *Zoological Journal of the Linnean Society*, **178**, 128–148.
 1194Salichos L., Rokas A. (2013) Inferring ancient divergences requires genes with strong
 1195phylogenetic signals. *Nature*, **497**, 327–331.
 1196Santos B. (2017) Phylogeny and reclassification of Cryptini (Hymenoptera, Ichneumonidae,
 1197Cryptinae), with implications for ichneumonid higher-level classification. *Systematic*
 1198*Entomology*, **42**, 650–676.
 1199Sharma P.P., Kaluziak S.T., Pérez-Porro A.R., Gonzáles V.L., Hormiga G., Wheeler W.C.,
 1200Giribet G. (2014) Phylogenomic interrogation of Arachnida reveals systemic conflicts in
 1201phylogenetic signal. *Molecular Biology and Evolution*, **31**, 2963–2984.
 1202Sheng M.-L., Broad G.R., Sun S.-P. (2012) A new genus and species of Collyriinae
 1203(Hymenoptera, Ichneumonidae). *Journal of Hymenoptera Research*, **23**, 103–125.
 1204Slater G.S., Birney E. (2005) Automated generation of heuristics for biological sequence
 1205comparison. *BMC Bioinformatics*, **6**, 31.
 1206Song S., Liu L., Edwards S.V., Wu S. (2012) Resolving conflict in the eutherian mammal
 1207phylogeny using phylogenomics and the multispecies coalescent model. *Proceedings of the*
 1208*National Academy of Sciences of the United States of America*, **109**, 14942–14947.
 1209Spasojevic T., Broad G.R., Bennett A.M.R., Klopstein S. (2018a) Ichneumonid parasitoid
 1210wasps from the Early Eocene Green River Formation: five new species and a revision of the
 1211known fauna (Hymenoptera, Ichneumonidae). *Paläontologische Zeitschrift*, **92**, 35–63.
 1212Spasojevic T., Wedmann S., Klopstein S. (2018b) Seven remarkable new fossil species of
 1213parasitoid wasps (Hymenoptera, Ichneumonidae) from the Eocene Messel Pit. *PLoS One*,
 1214**13**, e0197477.
 1215Stamatakis A. (2006) Phylogenetic models of rate heterogeneity: A high performance
 1216computing perspective. Proceedings of IPDPS2006. Rhodos, Greece.
 1217Stamatakis A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of
 1218large phylogenies. *Bioinformatics*, **30**, 1312–1313.
 1219Streicher J.W., Schulte J.A.I., Wiens J.J. (2016) How should genes and taxa be sampled for
 1220phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic*
 1221*Biology*, **65**, 128–145.
 1222Strimmer K., von Haeseler A. (1997) Likelihood-mapping: A simple method to visualize
 1223phylogenetic content of a sequence alignment. *Proceedings of the National Academy of*
 1224*Sciences of the United States of America*, **94**, 6815–6819.
 1225Suyama M., Torrents D., Bork P. (2006) PAL2NAL: robust conversion of protein sequence
 1226alignments into the corresponding codon alignments. *Nucleic Acids Research*, **34**, W609–
 1227W612.
 1228Swofford D.L. (2002) PAUP*. Phylogenetic analysis using parsimony (*and other methods).
 1229Version 4. Sinauer Associates, Sunderland, Massachusetts.
 1230Townes H.K. (1957) A revision of the genera of Poemeniini and Xoridini (Hymenoptera,
 1231Ichneumonidae). *Proceedings of the Entomological Society of Washington*, **59**, 15–23.
 1232Townes H.K. (1969) The genera of Ichneumonidae, Part 1. *Memoirs of the American*
 1233*Entomological Institute*, **11**, 1–300.
 1234Townes H.K. (1971) The genera of Ichneumonidae, Part 4. *Memoirs of the American*
 1235*Entomological Institute*, **17**, 1–372.
 1236Townsend J.P. (2007) Profiling phylogenetic informativeness. *Systematic Biology*, **56**, 222–
 1237231.
 1238Townsend J.P., Su Z., Tekle Y.I. (2012) Phylogenetic signal and noise: Predicting the power
 1239of a dataset to resolve phylogeny. *Systematic Biology*, **61**, 835–849.
 1240Virgilio M., Backeljau T., Nevado B., De Meyer M. (2010) Comparative performances of DNA
 1241barcoding across insect orders. *BMC Bioinformatics*, **11**, 206.

1242Waegele J.W., Mayer C. (2007) Visualizing differences in phylogenetic information content of
1243alignments and distinction of three classes of long-branch effects. *BMC Evolutionary Biology*,
1244**7**, No. 147.

1245Wahl D.B. (1986) Larval structures of oxytorines and their significance for the higher
1246classification of some Ichneumonidae (Hymenoptera). *Systematic Entomology*, **11**, 117–127.

1247Wahl D.B. (1990) A review of the mature larvae of Diplazontinae, with notes on larvae of
1248Acaenitinae and Orthocentrinae and proposal of two new subfamilies (Insecta: Hymenoptera,
1249Ichneumonidae). *Journal of Natural History*, **24**, 27–52.

1250Wahl D.B., Gauld I.D. (1998) The cladistics and higher classification of the Pimpliformes
1251(Hymenoptera : Ichneumonidae). *Systematic Entomology*, **23**, 265–298.

1252Waterhouse R.M., Tegenfeldt F., Li J., Zdobnov E.M., Kriventseva E.V. (2013) OrthoDB: a
1253hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Research*, **41**,
1254D358–D365.

1255Werren J.H., Richards S., Desjardins C.A., Niehuis O., Gadau J., Colbourne J.K., Group
1256T.N.G.W. (2010) Functional and evolutionary insights from the genomes of three parasitoid
1257*Nasonia* species. *Science*, **327**, 334–348.

1258Whitfield J.B., Austin A.D., Fernandez-Triana J. (2018) Systematics, biology, and evolution of
1259microgastrine parasitoid wasps. *Annual Review of Entomology*, **63**, 389–406.

1260Yang Z. (1996) Among-site rate variation and its impact on phylogenetic analyses. *Trends in*
1261*Ecology & Evolution*, **11**, 367–372.

1262Yang Z. (1998) On the best evolutionary rate for phylogenetic analysis. *Systematic Biology*,
1263**47**, 125–133.

1264Yu D.S., Van Achterberg C., Horstmann K. (2016) Taxapad 2016. Ichneumonoidea 2015
1265(Biological and taxonomical information), Taxapad Interactive Catalogue Database on flash-
1266drive. Nepean, Ottawa, Canada, www.taxapad.com.

1267Zaldivar-Riverón A., Shaw M.R., Sáez A.G., Mori M., Belokobylskij S.A., Shaw S.R., Quicke
1268D.L.J. (2008) Evolution of the parasitic wasp subfamily Rogadinae (Braconidae): phylogeny
1269and evolution of lepidopteran host ranges and mummy characteristics. *BMC Evolutionary*
1270*Biology*, **8**.

1271Zerbino D.R., Birney E. (2008) Velvet: algorithms for de novo short read assembly using de
1272Bruijn graphs *Genome Research*, **18**, 821–829.

1274**Tables**

1275Table 1. Taxon sampling for transcriptomics and hybrid enrichment.

Higher grouping	(Sub)family	Genus	Species	Specimen ID	Data type ¹	Country ²	Repos ³
Outgroup: Chalcidoidea	Pteromalidae	<i>cf. Eupelmophotismus</i>	sp.	Ichn-2467	hybEnr	AU	WC
Outgroup: Evanioidea	Gasteruptiidae	<i>Gasteruption</i>	sp.	Ichn-2470	hybEnr	AU	WC
Outgroup: Braconidae	Alysiinae	<i>Dacnusa</i>	<i>sibirica</i>	1KITE	transcr	DE	-
Outgroup: Braconidae	Aphidiinae	<i>Aphidius</i>	<i>colemanni</i>	1KITE	transcr	DE	-
Outgroup: Braconidae	Aphidiinae	<i>Diaeretus</i>	<i>essigellae</i>	1KITE	transcr	AU	-
Outgroup: Braconidae	Helconinae	Gen.	sp.	Ichn-2469	hybEnr	AU	-
Outgroup: Braconidae	Macrocentrinae	<i>Macrocentrus</i>	<i>marginator</i>	1KITE	transcr	SE	-
Outgroup: Braconidae	Microgastrinae	<i>Cotesia</i>	<i>vestalis</i>	1KITE	transcr	NL	-
Outgroup: Braconidae	Rogadinae	<i>Aleiodes</i>	<i>testaceus</i>	1KITE	transcr	FR	-
Ophioniformes	Anomaloninae	<i>Heteropelma</i>	<i>amictum</i>	1KITE	transcr	CH	-
Ophioniformes	Banchinae	<i>Apophua</i>	<i>evanescens</i>	transcr-SK09	transcr	CH	NMBE
Ophioniformes	Campopleginae	<i>Campoplex</i>	<i>capitator</i>	Ichn-2519	hybEnr	FR	NMBE
Ophioniformes	Campopleginae	<i>Hyposoter</i>	<i>didymator</i>	1KITE	transcr	FR	-
Ophioniformes	Cremastinae	<i>Dimophora</i>	<i>kentmartini</i>	Ichn-2465	hybEnr	AU	WC
Ophioniformes	Ctenopelmatinae	<i>Xenoschesis</i>	<i>fulvicornis</i>	Ichn-2359	hybEnr	CH	NMBE
Ophioniformes	Mesochorinae	<i>Astiphromma</i>	sp.	Ichn-2453	hybEnr	US	KYU
Ophioniformes	Metopiinae	<i>Colpotrochia</i>	<i>cincta</i>	transcr-SK10	transcr	CH	-
Ophioniformes	Ophioninae	<i>Leptophion</i>	<i>anici</i>	Ichn-2462	hybEnr	AU	WC
Ophioniformes	Tersilochinae	<i>Diaparsis</i>	sp.	Ichn-2449	hybEnr	CH	NMBE
Ophioniformes	Tryphoninae	<i>Netelia</i>	<i>cf. melanura</i>	1KITE	transcr	SE	-
Ophioniformes	Tryphoninae	<i>Netelia</i>	sp.	1KITE	transcr	HN	-
Ichneumoniformes	Alomyinae	<i>Alomya</i>	<i>debellator</i>	Ichn-2387	hybEnr	CH	NMBE
Ichneumoniformes	Cryptinae	<i>Buathra</i>	<i>laborator</i>	1KITE	transcr	SE	NMBE
Ichneumoniformes	Ichneumoninae	<i>Ichneumon</i>	<i>albiger</i>	1KITE	transcr	CH	-

Labeniformes	Labeninae	<i>Poecilocryptus</i>	cf. <i>nigromaculatus</i>	Ichn-2466	hybEnr	AU	WC
Pimpliformes	Acaenitinae	<i>Arotes</i>	<i>moiwanus</i>	Ichn-2341	hybEnr	JP	NMBE
Pimpliformes	Acaenitinae	<i>Coleocentrus</i>	<i>excitator</i>	transcr-SK01	transcr	CH	NMBE
Pimpliformes	Acaenitinae	<i>Jezarotes</i>	<i>tamanukii</i>	Ichn-2338	hybEnr	JP	NMBE
Pimpliformes	Acaenitinae	<i>Spilopteron</i>	<i>occiputale</i>	Ichn-2507	hybEnr	US	NMBE
Pimpliformes	Acaenitinae	<i>Yezoceryx</i>	sp.	Ichn-2509	hybEnr	TH	KYU
Pimpliformes	Collyriinae	<i>Collyria</i>	<i>trichophthalma</i>	transcr-SK11	transcr	CH	-
Pimpliformes	Cylloceriinae	<i>Cylloceria</i>	<i>melancholica</i>	Ichn-2131	hybEnr	SE	NMBE
Pimpliformes	Cylloceriinae	<i>Cylloceria</i>	cf. <i>caligata</i>	Ichn-2524	hybEnr	US	KYU
Pimpliformes	Diacritinae	<i>Diacritus</i>	<i>aciculatus</i>	Ichn-2143	hybEnr	SE	NMBE
Pimpliformes	Diacritinae	<i>Diacritus</i>	<i>aciculatus</i>	Ichn-2522	hybEnr	SE	NMBE
Pimpliformes	Diplazontinae	<i>Homotropus</i>	<i>frontorius</i>	Ichn-2365	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	<i>Sussaba</i>	<i>cognata</i>	Ichn-2513	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	<i>Syphophilus</i>	<i>tricinctorius</i>	1KITE	transcr	CH	-
Pimpliformes	Diplazontinae	<i>Syrphophilus</i>	<i>tricinctorius</i>	Ichn-2364	hybEnr	CH	NMBE
Pimpliformes	Diplazontinae	<i>Tymmophorus</i>	<i>suspiciosus</i>	Ichn-2460	hybEnr	CH	NMBE
Pimpliformes	Orthocentrinae	<i>Eusterinx</i>	<i>oligomera</i>	Ichn-2502	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Hemiphanes</i>	<i>erratum</i>	Ichn-2523	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Hemiphanes</i>	<i>performidatum</i>	Ichn-2499	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Megastylus</i>	<i>flavopictus</i>	Ichn-2183	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Neurateles</i>	sp.	Ichn-2526	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Orthocentrus</i>	sp.	Ichn-2500	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Picrostigeus</i>	<i>debilis</i>	Ichn-2498	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Plectiscidea</i>	cf. <i>tenuicornis</i>	Ichn-2503	hybEnr	SE	NMBE
Pimpliformes	Orthocentrinae	<i>Proclitus</i>	nr. <i>praetor</i>	Ichn-2510	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	<i>Acrodactyla</i>	<i>degener</i>	Ichn-2137	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	<i>Amazopimpla</i>	sp.	Ichn-2455	hybEnr	EC	UTU
Pimpliformes	Pimplinae	<i>Anastelgis</i>	sp.	Ichn-2454	hybEnr	NI	UTU
Pimpliformes	Pimplinae	<i>Apechthis</i>	<i>compunctor</i>	transcr-SK05	transcr	CH	NMBE
Pimpliformes	Pimplinae	<i>Atractogaster</i>	<i>semisculptus</i>	Ichn-2278	hybEnr	SE	SMTP

Pimpliformes	Pimplinae	<i>Camptotypus</i>	sp.	Ichn-2434	hybEnr	CG	NMBE
Pimpliformes	Pimplinae	<i>Delomerista</i>	<i>mandibularis</i>	Ichn-2528	hybEnr	SE	KYU
Pimpliformes	Pimplinae	<i>Dolichomitus</i>	<i>quercicolus</i>	transcr-SK04	transcr	CH	NMBE
Pimpliformes	Pimplinae	<i>Dolichomitus</i>	<i>imperator</i>	transcr-SK15	transcr	CH	-
Pimpliformes	Pimplinae	<i>Echthromorpha</i>	<i>intricatoria</i>	Ichn-2464	hybEnr	AU	WC
Pimpliformes	Pimplinae	<i>Flavopimpla</i>	<i>nigromaculata</i>	Ichn-2458	hybEnr	TH	KYU
Pimpliformes	Pimplinae	<i>Liotryphon</i>	cf. <i>punctulatus</i>	Ichn-2361	hybEnr	CH	NMBE
Pimpliformes	Pimplinae	<i>Nomosphaecia</i>	<i>pyramida</i>	Ichn-2093	hybEnr	TH	KYU
Pimpliformes	Pimplinae	<i>Parema</i>	<i>nigrobalteata</i>	Ichn-2497	hybEnr	TH	KYU
Pimpliformes	Pimplinae	<i>Perithous</i>	<i>scurra</i>	Ichn-2525	hybEnr	SE	SMTP
Pimpliformes	Pimplinae	<i>Pimpla</i>	<i>rufipes</i>	transcr-SK03	transcr	CH	-
Pimpliformes	Pimplinae	<i>Pimpla</i>	<i>flavicoxis</i>	1KITE	transcr	SE	NMBE
Pimpliformes	Pimplinae	<i>Piogaster</i>	<i>pilosator</i>	Ichn-2144	hybEnr	CH	NMBE
Pimpliformes	Pimplinae	<i>Polysphincta</i>	<i>tuberosa</i>	Ichn-2518	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	<i>Reclinervellus</i>	<i>nielsenii</i>	Ichn-2505	hybEnr	CZ	NMBE
Pimpliformes	Pimplinae	<i>Schizopyga</i>	<i>frigida</i>	Ichn-2136	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	<i>Theronia</i>	<i>hilaris</i>	Ichn-2433	hybEnr	US	NMBE
Pimpliformes	Pimplinae	<i>Xanthopimpla</i>	<i>varimaculata</i>	Ichn-2448	hybEnr	TH	KYU
Pimpliformes	Pimplinae	<i>Zatypota</i>	<i>percontatoria</i>	Ichn-2139	hybEnr	SE	NMBE
Pimpliformes	Pimplinae	<i>Zonopimpla</i>	sp.	Ichn-2457	hybEnr	GU	UTU
Pimpliformes	Poemeniinae	<i>Cnastis</i>	<i>vulgaris</i>	Ichn-2354	hybEnr	JP	NMBE
Pimpliformes	Poemeniinae	<i>Deuteroxorides</i>	<i>elevator</i>	transcr-SK07	transcr	CH	NMBE
Pimpliformes	Poemeniinae	<i>Eugalta</i>	cf. <i>strigosa</i>	Ichn-2442	hybEnr	TH	KYU
Pimpliformes	Poemeniinae	<i>Poemenia</i>	<i>hectica</i>	Ichn-2234	hybEnr	SE	NMBE
Pimpliformes	Poemeniinae	<i>Pseudorhyssa</i>	<i>nigricornis</i>	Ichn-2344	hybEnr	JP	NMBE
Pimpliformes	Rhyssinae	<i>Epirhyssa</i>	sp.	Ichn-2074	hybEnr	TH	KYU
Pimpliformes	Rhyssinae	<i>Rhyssa</i>	<i>amoena</i>	transcr-SK08	transcr	CH	-
Pimpliformes	Rhyssinae	<i>Triancyra</i>	<i>galloisi</i>	Ichn-2343	hybEnr	JP	NMBE
Xoridiformes	Xoridinae	<i>Xorides</i>	<i>praecatorius</i>	transcr-SK12	transcr	CH	NMBE

1277¹ Data type: transcr = transcriptome, hybEnr = hybrid enrichment data

1278² Country of origin of the specimen in ISO 3166-1 alpha-2 codes

1279³ Repositories of voucher specimens: KYU = Department of Entomology, University of Kentucky, USA; NMBE = Natural History Museum Bern,
1280Switzerland; SMTP = Swedish Malaise Trap Project, Station Linné on Öland, Sweden; UTU = University of Turku, Finland; WC = Waite Insect
1281collection, University of Adelaide, Australia. A dash indicates that no voucher specimens are available as they were ground for RNA extractions (if
1282a museum collection is mentioned for specimens used for transcriptome sequencing, it refers to a pseudo-voucher/para-or syngenophore, i.e.,
1283identical morphospecies collected at the same locality as the specimen whose transcriptome was sequenced).

1284

Table 2. Established oligonucleotide primer pairs for Sanger sequencing of eight exons. Gene numbers correspond to the *Nasonia* genome annotation and are preceded by “Nasvi2EG” on Genbank.

Gene	Primer pair	Primer sequence 5'–3'	Annealing temp	Fragment length
014339	F	GGC AAG AGG GTC AAA TTA CA	51–52 °C	142 bp + intron +190 bp
	R	CTC CAC AAT CTT TCC ATG CC		
011390	F	GAT CGT CGA AGT TAC AGC TA	49–50 °C	357 bp
	R	GGA GTG CCG GTC ATA AAG TC		
013087	F1	CTG AAG ACC ATT TCC CTG CG	51–52 °C	389 bp
	R1	GCG ACC TTG GAA GCA TCT TG		
	F2	CAG CAG CAA CGT GGA ATG G	51 °C	413 bp
	R2	GCG ACC TTG GAA GCA TCT TG		
001408	F	ACC GGT GGT CAC GAT TC	51 °C	377 bp
	R	CCA CAA TTT GAA TGT TCG ATC		
003698	F	CGA GCA GAG AGA GTA CAT G	51–52 °C	536 bp
	R	GGA AGA AGT CGT ATT TGA GGA A		
010812	F	GAT TAT CGT CCG TAC TTC AC	48–49 °C	140 bp + intron + 205 bp
	R	TAT CTC TCG AGG GGA ATC AT		
008805	F	GAG GAT GTC TCG AGC GC	52–53 °C	328 bp
	R	CTA GCG AAA TCA TAA GCG TAT C		
021461	F	GGA GTC GAC ATA AAC GGA TA	50–51 °C	404 bp
	R	GTG ATG TAG CCG TGT CTG AT		

1288

1289

1290

1291

1292

1293

Table 3. Summary of gene choice criteria.

Criterion ¹		Mean / median	95% range	#genes ²
Stationarity	p > 0.05	0.365 / 0.212	0.000–0.999	449
Clocklikeness	igrvar < 0.015	0.045 / 0.017	0.004–0.067	305

Root correct	bootstrap > 0.5	0.52 / 0.52	0.04–0.90	380
#SiteRates	> 80	117.7 / 115.5	73–178	683
Taxonomic bootscore	for final ranking	259.2 / 263.5	137–369	—

1295¹Criteria: Stationarity = homogeneity of nucleotide composition across taxa, tested by Chi-square test, with
1296genes with p-values above 0.05 retained. Clocklikeness = variance of IGR relaxed-clock model in single-gene
1297analysis, genes retained above 0.015. Root correct = bootstrap support for placement of *Xorides* as sister to all
1298other Ichneumonidae, gene retained if it was above 50%. #SiteRates = number of sites evolving at near-optimal
1299rate (according to Klopfstein *et al.*, 2017: between 0.1 and 0.4 substitutions from root to tip), genes retained if
1300they had more than 80. Taxonomic bootscore = sum of bootstrap support for 12 groupings chosen as unequivocal
1301in both our and previous analyses.

1302²Number of genes that would have been retained under each of the criteria if applied individually.

1303

1304Table 4. Number of transitions between idiobiont and koinobiont parasitoid strategy as
1305inferred from trees from the Bayesian posterior sample or from the ML bootstrap analyses,
1306along with the average of the estimated probabilities that the ancestor of Pimpliformes was an
1307idiobiont.

Analysis	Number of transitions				Probability (MRCA=idiobiont)
	media			max	
	mean	n	min		
PhyloBayes (AA)	5.0	5	5	6	0.99
MrBayes (AA)	5.8	6	5	6	0.18
RAxML (AA)	5.5	6	5	7	0.58
RAxML (12)	5.2	5	5	7	0.98
RAxML (123)	5.1	5	5	6	0.99

1308

1309

1310Figure legends

1311 Figure 1. Maximum likelihood trees from concatenated sequences of 723 genes
1312present in all transcriptome taxa. Support values are given as percentages, with maximum
1313support indicated by an asterisk. The tree on the left is the result of the analysis of partitioned
1314amino acid sequences, with support values representing non-parametric bootstraps and 10-
1315gene jackknifing. The tree on the right resulted from the analysis of the nucleotide data,
1316including all three codon positions. Support values represent bootstrap support from all three
1317codon positions, first and second, and percentages of the 10-gene jackknifing datasets
1318recovering the nodes, respectively. The only difference in the tree based on first and second
1319codon positions was the placement of *Rhyssa* as sister to ((*Collyria* + *Coleocentrus*) +
1320*Dolichomitus*).

1321 Figure 2. Consensus tree resulting from the Bayesian analysis of the amino acid
1322sequence data under an empirical profile mixture model as implemented in PhyloBayes, with
1323support values added from the partitioned Bayesian analysis in MrBayes, maximum
1324likelihood analysis in RAxML, and of the nucleotide analyses in RAxML. Support values
1325represent Bayesian posterior probabilities and bootstrap supports, respectively. Note the very
1326short branch lengths at the base of the Pimpliformes. Of taxa with an asterisk after the name,
1327we have analysed the transcriptomes, *Nasonia vitripennis* was included based on its published
1328data. Current subfamily classification of the pimpliform genera is indicated by coloured taxon
1329labels, which start with three letters corresponding to the subfamily name or to the family
1330(upper-case letters) name in the outgroups as follows: Aca = Acaenitinae, Alo = Alomyinae,
1331Ano = Anomaloninae, Ban = Banchinae, BRA = Braconidae, Cam = Campopleginae, Col =
1332Collyriinae, Cre = Cremastinae, Cry = Cryptinae, Cte = Ctenopelmatinae, Cyl =
1333Cylloceriinae, Dia = Diacritinae, Dip = Diplazontinae, GAS = Gasteruptiidae, Ich =
1334Ichneumoninae, Lab = Labeninae, Mes = Mesochorinae, Met = Metopiinae, Oph =
1335Ophioninae, Ort = Orthocentrinae, Pim = Pimplinae, Poe = Poemeniinae, PTE =
1336Pteromalidae, Rhy = Rhyssinae, Ter = Tersilochinae, Try = Tryphoninae, Xor = Xoridinae.

1337

1338 Figure 3. Phylogenetic uncertainty at the base of Pimpliformes shown by comparison
1339of consensus phylogenies from the Bayesian analyses (a, c), and maximum likelihood trees
1340from the ML analyses (d, e, f). Support values in the form of posterior probabilities (a, c) or
1341bootstrap percentages (d, e, f) are only shown if lower than 1.0 or 100%, respectively.

1342 Ancestral states of idiobiosis versus koinobiosis are shown as pies next to nodes, while stars
1343 indicate state transitions as inferred under maximum parsimony. Panel (b) shows one of the
1344 trees from the PhyloBayes analysis that shows an alternative reconstruction, with the
1345 reconstruction of the MRCA of Pimpliformes in favour of it being a koinobiont (arrow), as is
1346 also the case in panel (c). Outgroups and additional taxa within Ophioniformes and
1347 Ichneumoniformes were suppressed here for clarity of the figures.

1348

1349 Figure 4. Tree resolution as measured as the average bootstrap support in an ML
1350 analysis of all three codon positions plotted against the number of genes sampled (x-axis not
1351 in scale). Boxplots show the variation in twenty (for 2–30 genes) and 10 replicates (for 40–90
1352 genes), while lines represent the results from various analyses of genes sampled, non-
1353 randomly, after ranking according to three criteria.